IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 14 February 2024, accepted 29 February 2024, date of publication 1 April 2024, date of current version 24 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3383958

RESEARCH ARTICLE

Machine Learning Data-Driven Residential Load Multi-Level Forecasting With Univariate and Multivariate Time Series Models Toward Sustainable Smart Homes

LEILA ISMAIL[1,2], (Member, IEEE), HUNED MATERWALA[1,2], AND FIDA K. DANKAR[3]

[1]Intelligent Distributed Computing and Systems Lab, Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University, Al Ain, Abu Dhabi, United Arab Emirates
[2]Emirates Center for Mobility Research, United Arab Emirates University, Al Ain, Abu Dhabi, United Arab Emirates
[3]Computer Science Department, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Corresponding author: Leila Ismail (leila@uaeu.ac.ae)

This work was supported by the Emirates Center for Mobility Research, United Arab Emirates University, under Grant 12R199.

ABSTRACT Residential energy consumption is rapidly increasing every year due to demographic and behavioral changes, such as the rising population and the adoption of work-from-home post-COVID-19. High energy consumption emits a substantial amount of carbon dioxide and other Greenhouse Gases, contributing to global warming. It becomes crucial to accurately predict residential load. To enable smart home electricity consumption control, as well as efficient generation, planning, and usage, we predict household energy consumption at very short-term, short-term, and medium-term forecast levels using univariate and multivariate time series data. This study assesses the impact of different household units (water heater and air conditioning), areas (kitchen, laundry, office, living room, bathroom, ironing room, teenager room, and parents' room), and time (i.e., hour, day, and month) on energy consumption. Comparative analysis and numerical experimental results between the most used approaches, Support Vector Regression and Long Short-Term Memory, reveal that the former outperforms the latter across all forecast levels using different datasets. The findings of this paper will be useful to energy companies and household owners in enhancing energy efficiency and earning carbon credits by reducing the emission of carbon dioxide and other Greenhouse Gases.

INDEX TERMS Carbon credit, carbon emission, deep learning, energy consumption prediction, energy efficiency, forecast levels, Jensen-Shannon divergence, load forecasting, greenhouse gases (GHGs), long short-term memory (LSTM), machine learning, residential building energy consumption, root mean square error (RMSE), support vector regression (SVR), symmetric mean absolute percentage error (sMAPE), time series forecasting.

I. INTRODUCTION

A large proportion of global energy consumption comes from residential buildings. A report by the International Energy Agency shows that worldwide residential energy consumption has increased from 1203 Terawatt-hours (TWh) in 1974 to 6072 TWh in 2019, an increase of 404.74% [1]. In 2021, the residential sector accounted for 27% of the final energy consumption in the European Union, with the majority being used for space heating (i.e., 64.4%) followed by water heating (14.5%), lighting and electrical appliances (13.6%), cooking (6.0%), other uses (1.1%), and space cooling (0.5%) [2]. According to the Australian Government Department of Climate Change, Energy, the Environment, and Water, residential buildings in Australia account for 24% of overall electricity use and over 10% of the total carbon emission

The associate editor coordinating the review of this manuscript and approving it for publication was Christos Anagnostopoulos.

in the country [3]. Household electric energy demand and usage are significantly influenced by three factors: 1) environmental considerations (i.e., global warming and climatic change) [4], 2) technological and economic development (i.e., energy-efficient lighting and cooling equipment), and 3) social and demographic behaviors (i.e., work from home post-pandemic [5] and population growth [6]).

The daily household energy consumption is not constant and exhibits peaks and valleys depending on the time of day, day of the week, and month of the year [7]. These irregular demand curves do not align well with energy generation. Disproportionate generation of energy compared to the actual demand might result in blackouts if generated energy is less than the demand, or energy leaks if generated energy exceeds the demand [6], [7]. Furthermore, inefficient energy consumption leads to the accumulation of carbon dioxide and other Greenhouse Gases (GHGs) in the atmosphere, contributing to global warming and climate change [8]. To reduce global warming, the Kyoto Protocol [9] includes provisions that limit the emission of carbon dioxide and other GHGs using carbon credits, where one credit permits one ton of emission [10]. Businesses exceeding their carbon credit quotas should purchase extra credits for excess emissions, whereas those below can exchange or sell their credits [11].

Consequently, energy load forecasting becomes crucial for efficient energy generation and reducing carbon emissions. It aids the electricity-generating companies in predicting the required amount of energy to achieve a dynamic demand-supply equilibrium, plan for energy storage alternatives, and reduce energy costs [12]. In addition, forecasting at the sub-meter or appliances level can inform household owners about the areas of the house or appliances that will consume the most energy. This enables owners to proactively manage the use of areas or appliances in the home for energy efficiency [13]. Furthermore, accurate predictions will allow companies and household owners to earn carbon credits by improving energy efficiency.

Based on the time horizon of prediction, load forecasting can be classified into five different levels [14], [15], [16], [17]. (1) Very Short-Term Load Forecasting (VSTLF) focusing on a few minutes to an hour ahead prediction for production and management of daily electric load demand, as well as the prevention of blackouts. (2) Short-Term Load Forecasting (STLF) focusing on hourly, daily, or weekly ahead prediction for planning the production, transmission, and distribution of electric power. (3) Medium-Term Load Forecasting (MTLF) focusing on a few months to one year ahead prediction for planning major tests and maintenance schedules. (4) Long-Term Load Forecasting (LTLF) focusing on one year up to five years ahead prediction for national planning, investment, and the evaluation of energy contract prices. (5) Very Long-Term Load Forecasting (VLTLF) focusing on more than five years ahead prediction for scheduling the construction of new electric load generating units and planning environmental policies (such as looking up for renewable sources in case of high load forecast).

Energy consumption of a residential building can be forecasted either based on the previous timestamped energy consumption values (i.e., univariate time series) or based on other timestamped variables such as power, voltage, current, and weather conditions (i.e., multivariate time series) [18], [19]. The time series data is recorded over a fixed time interval (i.e., every minute, hour, day, week, month, year, etc.). Several works in literature have used different machine learning and deep learning approaches to forecast residential energy load by identifying energy consumption patterns in the time series data [6], [7], [8], [12], [13], [20], [21], [22], [23], [24], [25], [26], [27]. However, most of these works focus on a single forecast level [6], [7], [12], [13], [20], [21], [24], [25], [26], [27]. In this paper, we compare the performance of the two most used forecasting approaches in the literature, LSTM [28] and SVR [29], for three forecast levels, i.e., VSTLF, STLF, and MTLF. This is done by using publicly available Individual Household Electric Power Consumption (IHEPC) [30] and Appliances Energy Prediction (AEP) [21] datasets. The main contributions of this paper are as follows.

- We provide insights into the most consuming units and areas in the household based on the energy temporal distribution, trend, and seasonality.
- We evaluate the performance of the most used forecast approaches for VSTLF (every hour), STLF (every day and week), and MTLF (every month and quarter) forecast levels using the IHEPC dataset and for VSTLF (every 10 minutes and hour) and STLF (every day and week) forecast levels using the AEP dataset.
- We analyze the impact of global and household area-wise (i.e., kitchen, laundry, and heating and cooling) forecasts on the performance of learning algorithms under study with different forecast granularity for the IHEPC dataset.
- We analyze the impact of appliances' energy consumption forecast on the performance of learning algorithms with different forecast granularity for the AEP dataset.
- The performance of the algorithms is evaluated in terms of Symmetric Mean Absolute Percentage Error (sMAPE), Root Mean Square Error (RMSE), and Jensen-Shannon divergence.

The rest of the paper is organized as follows. Section II overviews the related work. The datasets, forecasting levels, and the learning algorithms employed in this study are described in Section III. Section IV presents the correlation among dataset features, their temporal dispersion, trends, and seasonality for each dataset. Numerical experiments and comparative performance results are provided in Section V. Finally, Section VI concludes the paper with future research directions.

TABLE 1. Work on residential energy load forecasting in the literature.

work	Dataset(s)	Forecast level					Algorithms	Evaluation metrics
		VSTLF	STLF	MTLF	LTLF	VLTLF		
[20]	Private dataset 1	×	×	✓ monthly	×	×	Average, FCNN, CNN, and LSTM	MAE, MAPE, and MdAPE
[21]	Appliances Energy Prediction (AEP) dataset	✓ 10 mins	×	×	×	×	MLR, SVR, RF, and GBM	RMSE, R-square, MAE, and MAPE
[6]	Smart Meters in London	✓ 30 mins	×	×	×	×	LSTM	MAE
[22]	**(1)** Net-Zero energy residential test facility dataset, **(2)** Individual household electric power consumption (IHEPC) dataset, and **(3)** UK DALE	✓ 24 hours	✓ 1 day and 7 days	×	×	×	RKF, Persistence, and LSTM	sMAPE
[23]	IHEPC dataset	✓ 15 mins and 1 hour	✓ 1 day and 1 week	×	✓ 1 year	×	CRBM, Factored CRBM, ANN, SVR, and RNN	RMSE and correlation coefficient R
[24]	UK Domestic Appliance Level Electricity (UK-DALE)	✓ 5, 10, 20, and 30 mins	×	×	×	×	SSA-PLSTM, DT, RF, MLP, SVR, LSTM, nested LSTM, EWT-LSTM, EMD-LSTM, VMD-LSTM, and SWT-LSTM	MAE, RMSE, MAPE, and R-square
[7]	Magicbox	✓ 15 mins	×	×	×	×	LSTM	MSE, RMSE, NRMSE, and Pearson coefficient
[12]	Private dataset 3	×	✓ Day	×	×	×	Naïve, LR, Prophet, and LSTM	MAPE
[8]	Customized dataset	×	×	×	×	×	NN	SSE, RMSE, R-square, and coefficient of variation
[25]	Private dataset 4	×	✓ 1 day	×	×	×	Transfer learning + LSTM (clustering), transfer learning + LSTM, and LSTM	RMSE, MAE, MAPE, learning and predicting times
[13]	Private dataset 5	✓ 15 mins	×	×	×	×	SARIMA + MetaFA-LSSVR, SARIMA, LSSVR, and MetaFA-LSSVR	R, RMSE, MAE, MAPE, MaxAE, TER, and CPU time
[26]	Customized dataset	✓ 1 hour	×	×	×	×	DNN, ANN, and DT	R-square, MAR, and model building and training times
[27]	IHEPC dataset	✓ 1 min	×	×	×	×	DNN, Fine tree regression, SVR, and ANN	MSE, RMSE, and MAPE
This work	IHEPC dataset	✓ hour	✓ day and week	✓ month and quarter	×	×	SVR and LSTM	sMAPE, RMSE, and Jensen-Shannon divergence
	AEP dataset	✓ 10 mins and hour	✓ day and week	×	×	×		

ANN: Artificial Neural Network; CNN: Convolutional Neural Network; CRBM: Conditional Restricted Boltzmann Machines; DNN: Deep Neural Network; DT: Decision tree; EMD: Empirical Mode Decomposition; EWT: Empirical Wavelet Transform; FCNN: Fully Connected Neural Network; GBM: Gradient Boosting Machine; LR: Linear Regression; LSSVR: Least Square Support Vector Regression; LSTM: Long Short-Term Memory; LTLF: Long Term Load Forecasting; MAE: Mean Absolute Error; MAPE: Mean Absolute Percentage Error; MAR: Mean Absolute Residual; MaxAE: Maximum Absolute Error; MdAPE: Median Absolute Percentage Error; MetaFA: Meta Firefly Algorithm; MLP: Multilayer Perceptron; MLR: Multiple Linear Regression; MSE: Mean Square Error; MTLF: Medium Term Load Forecasting; NN: Neural Network; NRMSE: Normalized RMSE; RF: Random Forest; RKF: Recontextualized Kalman Filter; RMSE: Root Mean Square Error; RNN: Recurrent Neural Network; SARIMA: Seasonal Autoregressive Integrated Moving Average; sMAPE: Symmetric MAPE; SSA-PLSTM: Singular Spectrum Analysis – Parallel LSTM; SSE: Sum Squared Error; STLF: Short Term Load Forecasting; SVR: Support Vector Regression; SWT: Stationary Wavelet Transform; TER: Total Error Rate; VLTLF: Very Long Term Load Forecasting; VMD: Variational Mode Decomposition; VSTLF: Very Short Term Load Forecasting.

II. RELATED WORK

Table 1 summarizes the works on machine learning and deep learning-based energy load forecasting [6], [7], [8], [12], [13], [20], [21], [22], [23], [24], [25], [26], [27]. It presents the dataset(s) used by these works, the forecast level (i.e., VSTLF, STLF, MTLF, LTLF, and VLTLF), the implemented algorithms, and the considered evaluation metrics. As stated in the table, most works perform a single forecast level [6], [7], [12], [13], [20], [21], [24], [25], [26], [27]. In particular, [6], [7], [13], [21], [24], [26], [27] focus on VSTLF, [12], [25] on STLF, and [20] on MTLF. In contrast, [22] implements different learning algorithms for VSTLF and STLF, whereas [23] evaluated the performance of learning algorithms for VSTLF, STLF, and LTLF. Table 2 describes the datasets used by the works in literature, along with the data collection period, number of records, and considered features. As presented in the table, works [8], [12], [13], [20], [25], [26] use a private/customized dataset for load forecasting, whereas [6], [7], [21], [22], [23], [24], [27] use publicly available datasets.

Based on Table 1, it is evident that LSTM and SVR are the most used learning algorithms for all forecast levels. However, these two algorithms are compared in a unified setup for only VSTLF [24]. Furthermore, the dataset used to evaluate these algorithms is relatively small (>10,000 records) compared to the other publicly available datasets (Table 2). In this paper, we address this void by comparing the performances of LSTM and SVR for VSTLF, STLF, and MTLF using publicly available IHEPC (with the highest number of records) and AEP (with the highest number of records including appliances' energy consumption) datasets.

III. METHODOLOGY

A. DATASETS

The Individual Household Electric Power Consumption (IHEPC) dataset [30] by Hebrail and Berard and the

TABLE 2. Datasets used for energy load forecasting in the literature.

Work	Dataset	Country	Description	Data interval	Number of records	Number of features	Features used
[20]	Private dataset 1	Brazil	Energy consumption data, collected every month, from Brazilian power company	03/2014 – 02/2016	21978437 measurements from 934945 customers	1	Energy consumption
[21]	Appliances Energy Prediction dataset [21]	Stambruges, Belgium	Energy consumption data for a house is collected every 10 minutes using electric energy metering. House temperature and humidity conditions are measured using a wireless sensor network. Weather data is obtained from the nearest airport weather station	137 days (4.5 months)	19735	31	Appliances energy consumption, light energy consumption, temperature in kitchen area, humidity in kitchen area, temperature in living room area, humidity in living room area, temperature in laundry room area, humidity in laundry room area, temperature in office room, humidity in office room, temperature in bathroom, humidity in bathroom, temperature outside the building (north side), humidity outside the building (north side), temperature in ironing room, humidity in ironing room, temperature in teenager room 2, humidity in teenager room 2, temperature in parents room, humidity in parents room, temperature outside (from weather station), pressure (from weather station), humidity outside (from weather station), windspeed (from weather station), visibility (from weather station), Tdewpoint (from weather station), random Variable 1, random Variable 2, number of seconds from midnight, week status (weekend (0) or a weekday (1)), day of week, and date time stamp
[6]	Smart Meters in London [31]	London	Energy readings from 5567 households at a sampling rate of 1 sample per 30 min	11/2011 – 02/2014	Not reported	16	Energy consumption and socioeconomic background of the residents. The socioeconomic features are categorized into population, housing, finance, transport, environment, leisure time, digital, marketing, health, contact, safety, education, shopping, family, and economy
[22]	Net-Zero energy residential test facility dataset [32]	Gaithersburg, U.S	Energy consumption data, sampled every minute, collected from a single housing testbed on the campus of the National Institute of Standards of Technology	01/02/2015 – 31/01/2016	Not reported	63	Energy consumption data of 63 appliances
[22], [23], [27]	Individual household electric power consumption dataset [30]	Sceaux, France	Energy consumption data of a household, collected every minute using one global meter and three submeters	17/12/2006 – 12/12/2010	2075259	9	Date (dd/mm/yyyy), time, Global active power, global reactive power, voltage, global intensity (current intensity), sub metering 1 (energy corresponding to the kitchen), sub metering 2 (energy corresponding to the laundry room), and sub metering 3 (energy corresponding to water-heater and air-conditioner)
[22]	UK Domestic Appliance Level Electricity (UK-DALE) [33]	London, United Kingdom	Energy consumption data from 5 residential households sampled every 6s	09/11/2012 – 26/04/2017	>10,000	53	Energy consumption data of 1 global meter and 52 appliances
[24]					>10,000	Not reported	Not reported
[7]	Magicbox [34]	Madrid	Energy consumption and weather-related data extracted every minute from a real solar house, Magicbox	07/2016 – 09/2019	> 1.5 M	7	Previous day energy consumption, outdoor temperature, relative humidity, irradiance, indoor CO2 level, indoor temperature, reference temperature
[12]	Private dataset 3	Silesian voivodeship	Energy consumption data, collected every few seconds from 4 non-commercial buildings that include a flat with a family of 2+1, a flat with 2 adults, a modern detached house with a family of 2+1, a detached house with a family of 2+2	01/03/2021 – 28/10/2021	Not reported	Not reported	Energy consumption of most energy-intensive devices and overall energy consumption of the whole location
[8]	Customized dataset	Canada	Energy consumption data from the 1993 survey of household energy use (SEHU) [35] and 1993 heating and cooling degree-day data	1993	8767 households	56	Boiler pump, central electronic air-filter, central electronic dehumidifier, central electronic humidifier, central vacuum-cleaner, central ventilation system, electrical cooking appliance, furnace fan, heat-recovery ventilation system, jacuzzi, kitchen exhaust-fan, microwave, sauna, sump pump, water cooler, water softener, portable dehumidifier, black-and-white TV, CD player, portable electric-heater, portable humidifier, bathroom exhaust-fan, computer, interior car warmer, VCR, water bed, stereo, car block heater, ceiling fan, color TV, electric blanket, fish tank, portable fan, clothes dryer, clothes washer, dishwasher, main refrigerator, second refrigerator, main freezer, second freezer, central A/C, window A/C, HDD, CDD, halogen lights, fluorescent lights, incandescent lights, total heated area , income ($10,000/year) , dwelling type (1 if single-detached; 0 if single-attached), dwelling ownership (1 if owner; 0 if renter), size of area of residence, employed adult ratio, number of employed adults / number of adults, number of children, number of adults

TABLE 2. *(Continued.)* **Datasets used for energy load forecasting in the literature.**

[25]	Private dataset 4	South Korea	15-min demand loads of several apartments, collected from two smart buildings	01/2016 – 12/2018	Not reported	1	Energy demands of apartments (profiles) in each building
[13]	Private dataset 5	New Taipei, Taiwan	Real-time energy consumption data, collected every minute from a smart grid of a three-story building with a family of 2+3	22/06/2015 – 26/07/2015	1440 data points each day	3	Energy usage of appliances 1st (electric fan, personal computer, desk lamp, wall lighting, cell lighting, specialized machines, and dehumidifier), 2nd(TV set, cell lighting, electric fan, oven, microwave, rice cook, hot water machine, water heater, stereo, refrigerator, washing machine, and small lighting), and 3rd (cell lighting, wall lighting, personal computer, lamp, air conditioning) floors and electrical devices
[26]	Customized dataset	Canada	Hourly power demand and energy prices in Ontario province in Canada, obtained from Independent Electricity System Operator, and weather conditions obtained from Canadian Climate Data – Environment Canada	Not reported	200	11	Hourly power load, last year's load at the same time, previous week's load for the same time, previous day's load for the same time, average load for 24 hours before this time, outside temperature, dew point temperature, real humidity, average windspeed, outside air pressure, and hourly energy prices.

TABLE 3. **Characteristics of the datasets used in the experiments.**

	IHEPC dataset	AEP dataset
Dataset description	Measurements collected from a house located in Sceaux (7km from Paris, France)	Energy measurements collected from a house located in Stambruges (about 24 km from the City of Mons, Belgium), and weather data obtained from the nearest airport weather station (Chièvres Airport, Belgium)
Date range	16 December 2006 to 26 November 2010	11 January 2016 to 27 May 2016
Sampling period	1 minute	10 minutes
Measurements	2075259 (before preprocessing) and 2049280 (after preprocessing)	19735
Features	9	29

TABLE 4. **IHEPC dataset features and their description.**

Feature	Description	Unit	Minimum value	Maximum value	Mean	Standard deviation
Date	The date on which the measurements are recorded	dd/mm/yyyy	16/12/2006	26/11/2010	Not applicable	Not applicable
Time	Time at which the measurements are recorded	hh:mm:ss	17:24:00	21:02:00	Not applicable	Not applicable
Global active power	Minute-averaged global active power of the household	kilowatt	0.076	11.122	1.092	1.057
Global reactive power	Minute-averaged global reactive, i.e., unused, power of the household	kilowatt	0	1.390	0.124	0.113
Voltage	Minute-averaged voltage	volt	223.200	254.150	240.840	3.240
Global intensity	Minute-averaged global current intensity of the house	ampere	0.200	48.400	4.628	4.444
Sub metering 1	Energy consumption corresponding to a kitchen consisting of a dishwasher, an oven, and a microwave	watt-hour	0	88	1.122	6.153
Sub metering 2	Energy consumption corresponding to laundry room consisting of a washing machine, a tumble-drier, a refrigerator, and a light	watt-hour	0	80	1.300	5.822
Sub metering 3	Energy consumption corresponding to an electric water heater and an air-conditioner	watt-hour	0	31	6.460	8.437

Note – The minimum, maximum, mean, and standard deviation decimal values are presented with precision to three decimal points

TABLE 5. AEP dataset features and their description.

Feature	Description	Unit	Minimum value	Maximum value	Mean	Standard deviation
Date	Timestamp of the measurement	dd/mm/yyyy hh:mm:ss	11/2/2016 17:00:00	27/05/2016 18:00:00	Not applicable	Not applicable
Appliances	Energy consumption of appliances	Watt-hour	10	1080	97.694	102.524
Lights	Energy consumption of light fixtures	Watt-hour	0	70	3.801	7.935
T1	Temperature in kitchen area	Celsius	16.790	26.260	21.686	1.606
RH_1	Humidity in kitchen area	%	27.023	63.600	40.259	3.979
T2	Temperature in living room area	Celsius	16.100	29.586	20.341	2.192
RH_2	Humidity in living room area	%	20.463	56.026	40.420	4.069
T3	Temperature in laundry area	Celsius	17.200	29.236	22.267	2.006
RH_3	Humidity in laundry area	%	28.766	50.163	39.242	3.254
T4	Temperature in office room	Celsius	15.100	26.200	20.855	2.042
RH_4	Humidity in office room	%	27.660	51.090	39.026	4.341
T5	Temperature in bathroom	Celsius	15.330	25.795	19.592	1.844
RH_5	Humidity in bathroom	%	29.815	96.321	50.949	9.022
T6	Temperature outside the building (north side)	Celsius	-6.065	28.290	7.910	6.090
RH_6	Humidity outside the building (north side)	%	1	99.900	54.609	31.149
T7	Temperature in ironing room	Celsius	15.390	26	20.267	2.109
RH_7	Humidity in ironing room	%	23.200	51.400	35.388	5.114
T8	Temperature in teenager room 2	Celsius	16.306	27.230	22.029	1.956
RH_8	Humidity in teenager room 2	%	29.600	58.780	42.936	5.224
T9	Temperature in parents' room	Celsius	14.890	24.500	19.485	2.014
RH_9	Humidity in parents' room	%	29.166	53.326	41.552	4.151
To	Temperature outside (from Chièvres weather station)	Celsius	-5	26.100	7.411	5.317
Pressure	Pressure (from Chièvres weather station)	mm Hg	729.300	772.300	755.522	7.399
RH_out	Humidity outside (from Chièvres weather station)	%	24	100	79.750	14.901
Windspeed	Speed of wind (from Chièvres weather station)	Meter/second	0	14	4.039	2.451
Visibility	Visibility (from Chièvres weather station)	Kilometer	1	66	38.330	11.794
Tdewpoint	Dew point temperature (from Chièvres weather station)	Celsius	-6.600	15.500	3.760	4.194
rv1	Random variable 1	Non-dimensional	0.005	49.996	24.998	14.496
rv2	Random variable 2	Non-dimensional	0.005	49.996	24.998	14.496

Note – The minimum, maximum, mean, and standard deviation decimal values are presented with precision to three decimal points

Appliances Energy Prediction (AEP) dataset [21] are used to evaluate the performance of electric load forecasting models. The IHEPC dataset is selected because it contains the highest number of records with sub metering energy consumption data, while the AEP dataset is selected because it contains the highest number of records with appliances' energy consumption data (Table 2). The characteristics of the datasets are described in Table 3. Tables 4 and 5 present the description and statistical information of the IHEPC and AEP datasets respectively. As depicted in the tables, both datasets are multivariate time series. Before preprocessing, each record in the IHEPC dataset consists of 9 features, whereas each record in the AEP dataset consists of 29 features. We preprocessed the IHEPC dataset by combining the 'Date' and 'Time' features into a 'Timestamp' and removing the records with missing values. In total, we removed 25,797, i.e., 1.25% of the total records. The AEP dataset has no missing values. The 'Sub metering 1', 'Sub metering 2', 'Sub metering 3', and 'Global active power' are predicted for the IHEPC dataset, whereas the 'Appliances' energy consumption is predicted for the AEP dataset.

B. FORECASTING LEVEL

The IHEPC dataset is used to forecast electricity load for three different forecast levels: 1) VSTLF, 2) STLF, and 3) MTLF. This is through downsampling the dataset by decreasing the frequency of the recordings. In particular, for VSTLF level, $IHEPC_{hour}$ dataset is created by downsampling the IHEPC dataset from minutes to hours sampling frequency. This is done by aggregating the recordings sampled every minute in the dataset for each hour. Similarly, for STLF, $IHEPC_{day}$ and $IHEPC_{week}$ datasets are created by downsampling the IHEPC dataset to days and weeks sampling frequencies respectively. For MTLF, $IHEPC_{month}$ and $IHEPC_{quarter}$

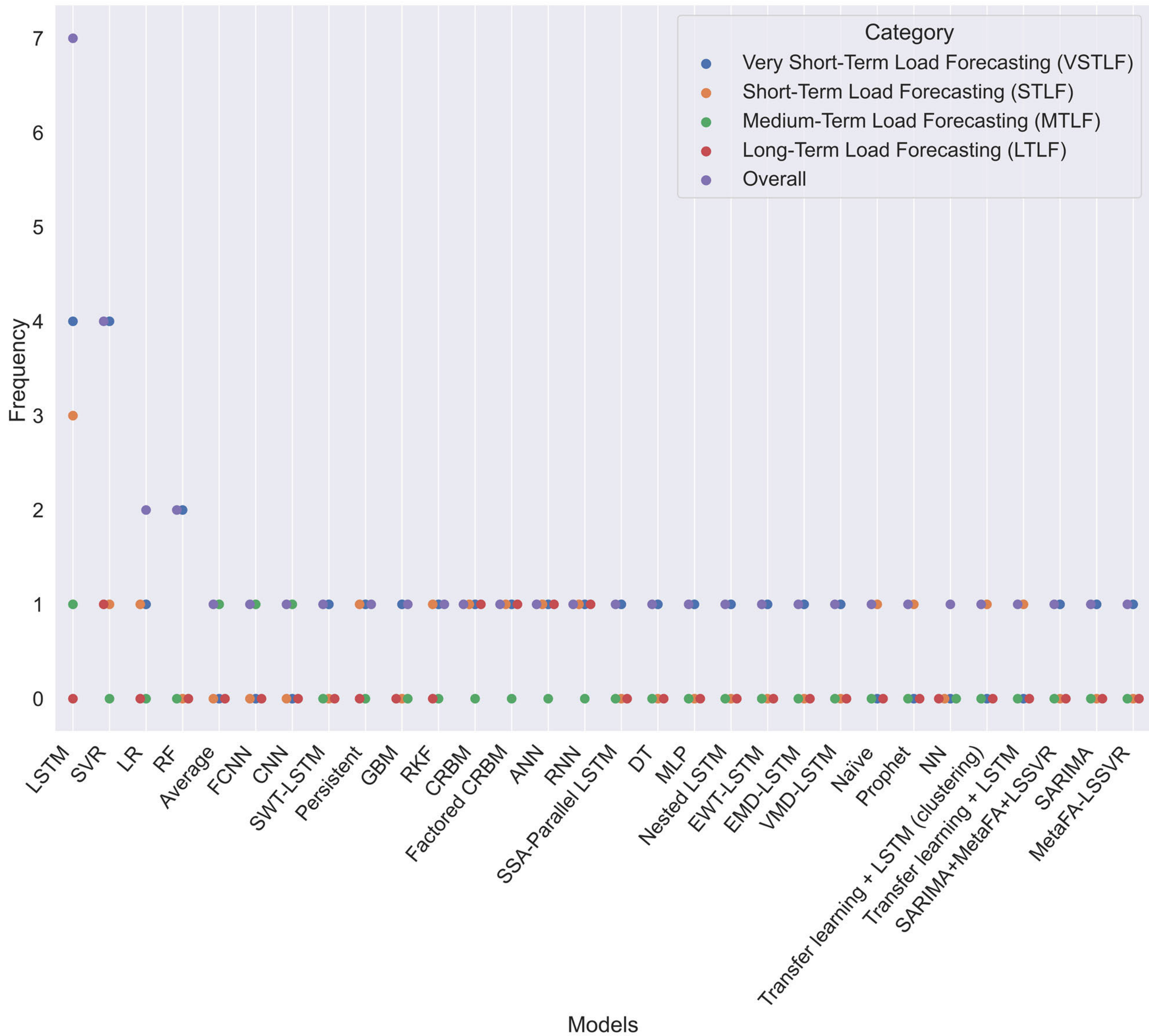

FIGURE 1. Frequency of learning algorithms used in literature for energy load forecasting (ANN: Artificial Neural Network; CNN: Convolutional Neural Network; CRBM: Conditional Restricted Boltzmann Machines; DT: Decision tree; EMD: Empirical Mode Decomposition; EWT: Empirical Wavelet Transform; FCNN: Fully Connected Neural Network; GBM: Gradient Boosting Machine; LR: Linear Regression; LSSVR: Least Square Support Vector Regression; LSTM: Long Short-Term Memory; MetaFA: Meta Firefly Algorithm; MLP: Multilayer Perceptron; MLR: Multiple Linear Regression; NN: Neural Network; RF: Random Forest; RKF: Recontextualized Kalman Filter; RNN: Recurrent Neural Network; SARIMA: Seasonal Autoregressive Integrated Moving Average; SSA-PLSTM: Singular Spectrum Analysis - Parallel LSTM; SVR: Support Vector Regression; SWT: Stationary Wavelet Transform; VMD: Variational Mode Decomposition.

datasets are created by downsampling the IHEPC dataset to months and quarters sampling frequencies respectively. The resulting $IHEPC_{hour}$, $IHEPC_{day}$, $IHEPC_{week}$, $IHEPC_{month}$, and $IHEPC_{quarter}$ datasets consist of 34589, 1442, 207, 48, and 17 records respectively. In this study, LTLF for the IHEPC dataset is not considered as downsampling the dataset to yearly sampling frequency would result in a dataset containing only 5 records, with 4 records allocated for training and 1 for validation. Having only 1 record for validation leads to unreliable performance evaluation of a forecasting model. Furthermore, VLTLF for the IHEPC dataset is not considered as the dataset collected over 4 years cannot be downsampled to a frequency greater than 5 years.

The AEP dataset is used to forecast appliances' energy for two forecast levels: 1) VSTLF and 2) STLF. VSTLF level involves both minute and hourly predictions. For minute predictions, the AEP dataset is used as it is, while for hourly predictions, AEP_{hour} dataset is created by downsampling the AEP dataset from minutes to hours sampling frequency. This is done by aggregating the recordings sampled every 10 minutes in the dataset for each hour. Similarly, for STLF, AEP_{day} and AEP_{week} datasets are created by downsampling the AEP dataset to days and weeks sampling frequencies respectively. The resulting AEP_{hour}, AEP_{day}, and AEP_{week} datasets consist of 3290, 138, and 20 records respectively. In this study, MTLF, LTLF, and VLTLF levels for the AEP dataset are not considered as the dataset collected over 4.5 months cannot be downsampled to quarterly or yearly frequencies.

C. ENERGY LOAD FORECASTING ALGORITHMS

In this section, we explain machine learning and deep algorithms under study for energy load forecasting. We evaluate the performance of LSTM and SVR algorithms as they are the most used approaches for VSTLF, STLF, MTLF, and LTLF in literature as shown in Figure 1.

1) LONG SHORT-TERM MEMORY (LSTM)

Long Short-Term Memory (LSTM) is a type of recurrent neural network model [28] that consists of gate structures and memory blocks to recognize temporal dependency between time-series load forecasting datasets. It solves the issue of vanishing or exploding gradient problems [36] while updating the weights to effectively learn long-term temporal structures. The main components of the LSTM network are the memory cell (also referred to as cell state) and the gates as depicted in Figure 2.

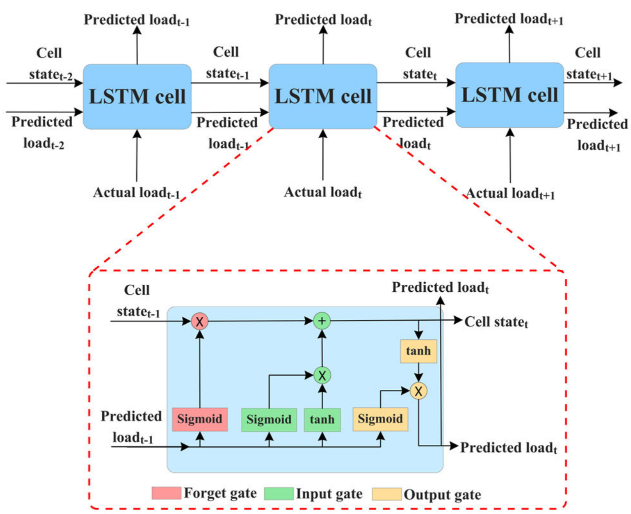

FIGURE 2. Structure of Long Short-Term Memory (LSTM) network.

The cell state transfers relevant information to the chain of neural networks while gates aid in removing or adding specific information to the internal state by controlling the update of the cell state. There are three different types of gates in an LSTM cell; 1) Forget gate that decides on which information should be removed, 2) Input gate that selects values from the input to update the memory state, and 3) Output gate that decides on the output or next hidden state value.

2) SUPPORT VECTOR REGRESSION (SVR)

Support Vector Regression (SVR) [29] is an extension of Support Vector Machine (SVM) to solve regression problems. It uses support vectors to define a hyperplane (i.e., the best-fit line) that includes the maximum number of recordings from the dataset. Support vectors are the recordings in the dataset that are closest to the hyperplane and removing them from the dataset will change the position of the hyperplane. For datasets with non-linearity, SVR uses a kernel function that transforms the dataset space into a higher dimension, without increasing the computational cost, to find a hyperplane. Different types of kernel functions are linear, polynomial, and Radial Basis Functions (RBF).

IV. DATA EXPLORATION

In this section, for each dataset under study, we analyze the correlations between the dataset features and examine the temporal distribution of these features for different time horizons. In addition, we decompose each feature to identify the underlying data trend and seasonality.

A. IHEPC DATASET

1) FEATURES CORRELATION

Figure 3 illustrates the Pearson correlation coefficients [37] between the features in the dataset under study. As depicted, global active power is tightly correlated with global current intensity. This relationship is intuitive, given that higher current intensity corresponds to increased power consumption. Compared to the kitchen area (Sub_metering_1) and the laundry area (Sub_metering_2), the energy consumption of the air conditioner and electric water heater (Sub_metering_3) exhibits a stronger correlation with global active power consumption. No systematic association is found between the energy consumption of the kitchen area, laundry area, and air conditioning and water heater.

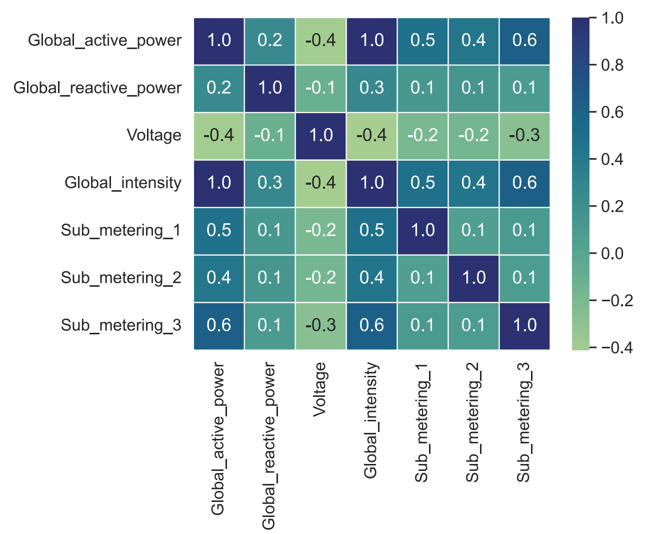

FIGURE 3. Pearson correlation between features in the IHEPC dataset.

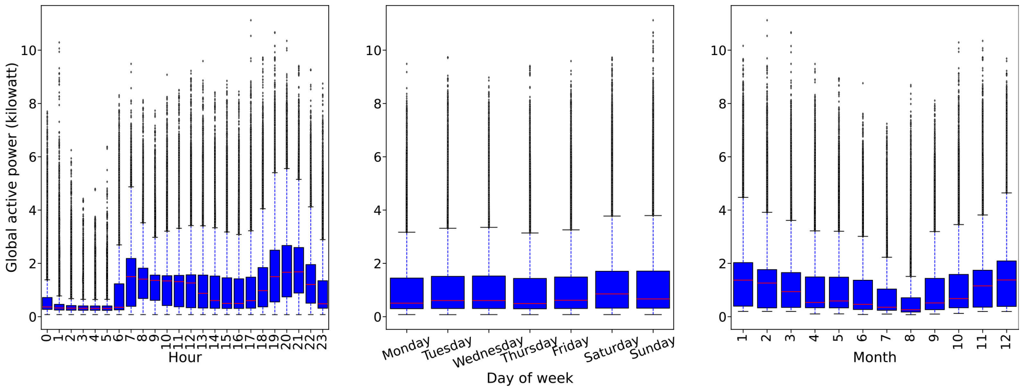

FIGURE 4. Hourly, weekly, and monthly variations in global active power.

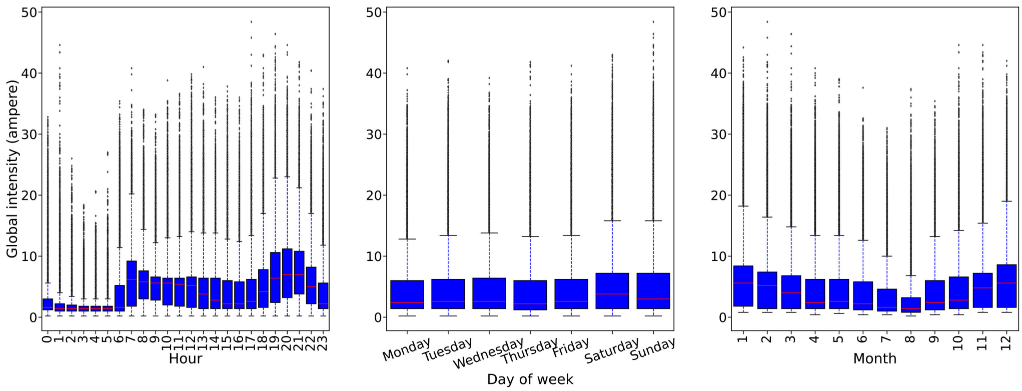

FIGURE 5. Hourly, weekly, and monthly variations in global intensity.

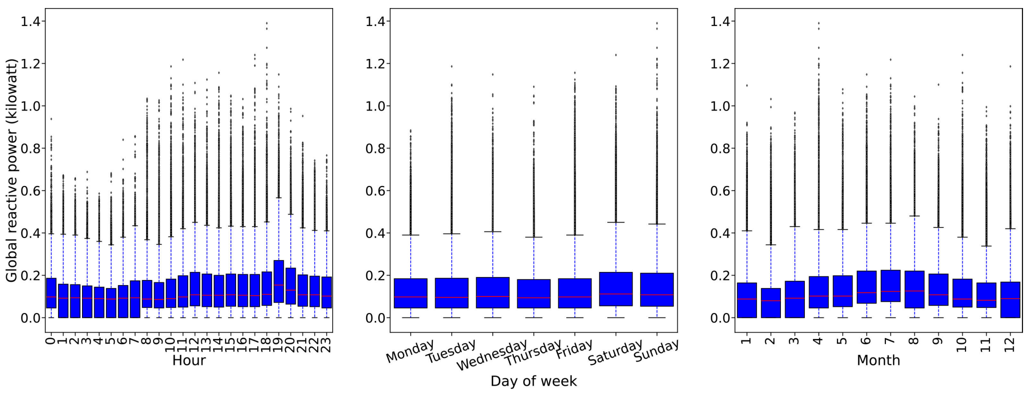

FIGURE 6. Hourly, weekly, and monthly variations in global reactive power.

2) TEMPORAL DISPERSION OF DATASET FEATURES

Figures 4 – 10 show the temporal variation of the features in the IHEPC dataset across three different time horizons: (i) intra-day variations over hours, (ii) weekly variations over days of the week, and (iii) seasonal variations over months. As depicted in Figure 4, the distribution of global active power consumption is positively skewed. Power consumption remains constant with the least median between 01:00 a.m. and 05:00 a.m., corresponding to the sleeping hours of the individuals in the house from which the data is collected. As the day begins, the power consumption slightly increases at 06:00 a.m. and spikes at 07:00 a.m., potentially due to the use of the kitchen area during breakfast. Throughout the day (office/school hours), power consumption gradually decreases until 03:00 p.m. As people return home from work or school, power consumption gradually increases after 04:00 p.m. with the maximum median at 08:00 p.m. and 09:00 p.m. (i.e., dinner time). Consumption then decreases

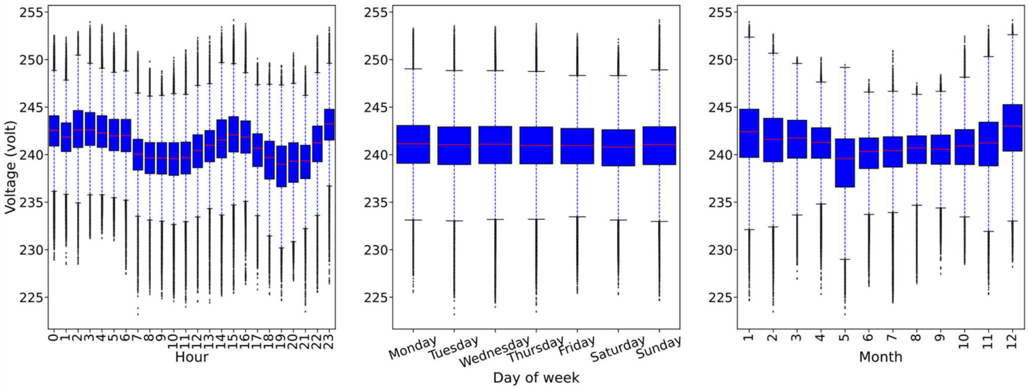

FIGURE 7. Hourly, weekly, and monthly variations in voltage.

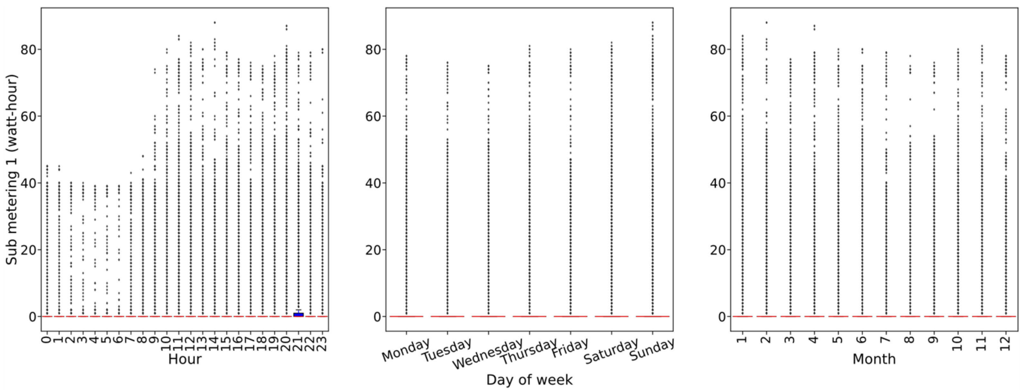

FIGURE 8. Hourly, weekly, and monthly variations in sub metering 1 (i.e., kitchen area).

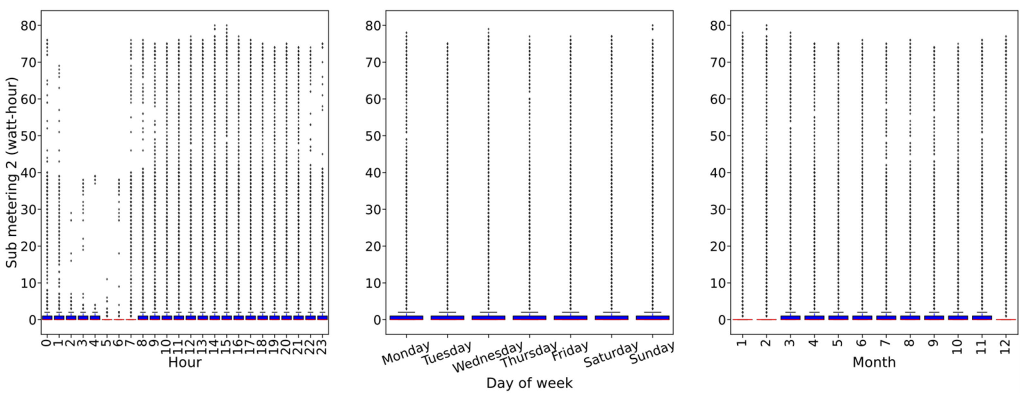

FIGURE 9. Hourly, weekly, and monthly variations in sub metering 2 (i.e., laundry area).

until midnight. Concerning the days of the week, the median power consumption is almost the same for all the days, with a slightly higher value on Saturdays, likely due to the weekend. On a seasonal basis, the median power consumption decreases from January, reaching its minimum in August, after which it increases. High power consumption in November, December, January, February, and March is due to the end of fall and the entire winter season in France, leading to increased use of water heaters. A similar temporal variation is observed for global intensity as shown in Figure 5. This is because global active power and global intensity are highly correlated (Figure 3).

As shown in Figure 6, the distribution of global reactive power consumption is positively skewed. The median global reactive power remains constant from midnight until 06:00 p.m., increases to the maximum at 07:00 p.m., and

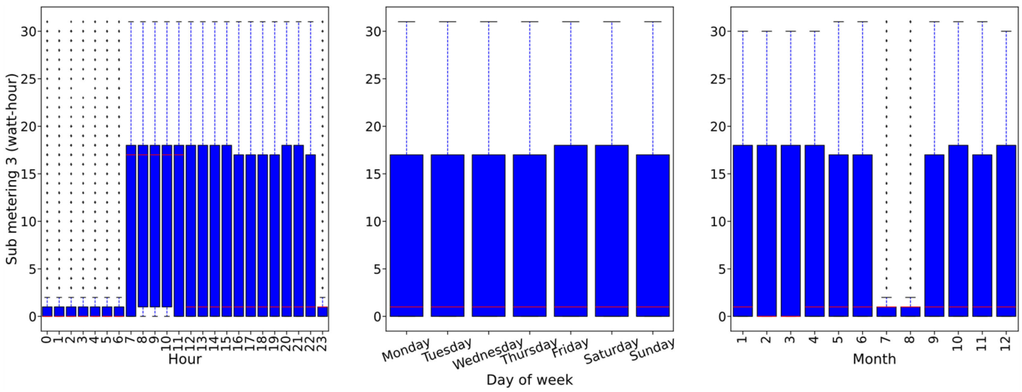

FIGURE 10. Hourly, weekly, and monthly variations in sub metering 3 (i.e., air conditioner and electric water heater).

then decreases until 11:00 p.m. It remains nearly constant throughout the week. The median global reactive power almost remains constant over months, with a slightly higher median in July, August, and September. Figure 7 reveals that voltage follows a normal distribution with almost equal whisker lengths on both sides of each box. The voltage maintains a constant median between midnight and 03:00 a.m., with a slight dip at 01:00 a.m. It slightly decreases until 05:00 a.m., remains steady at 06:00 a.m., and then drops at 7 a.m. during breakfast time. From 08:00 – 11:00 a.m., the median voltage remains constant, increases in the afternoon until 03:00 p.m., and then decreases to the lowest median at 07:00 p.m. The median voltage is the highest at 11:00 p.m. During the days of the week, the median voltage remains almost constant. Regarding seasonal variations, the median voltage is highest in December and January, lowest in May, and remains constant from June to September (i.e., summer season and the beginning of fall).

Figures 8 –10 reveal the positively skewed variation distributions for sub metering 1 (i.e., kitchen area), sub metering 2 (i.e., laundry area), and sub metering 3 (i.e., air conditioner and electric water heater) respectively. As shown in Figure 8, the median energy consumption for the kitchen area remains almost zero throughout the day. Moreover, the median for every hour, except at 09:00 p.m., coincides with the first and third quartiles of the box plot, indicating consistently low and identical energy consumption values. On the other hand, at 09:00 p.m., the median energy consumption aligns with the first quartile, revealing that a large proportion of energy consumption values is low and identical, with a few data points representing higher energy consumption. This might be due to the use of a dishwasher after dinner. The median energy consumption remains near zero throughout the week and over the months. Regarding the laundry area (Figure 9), a significant proportion of energy consumption data points are low and likely identical throughout the day, except from 05:00 a.m. – 07:00 a.m., where all data points are low and probably identical. The median energy consumption for sub metering 2 remains constant over the weeks and months. However, compared to the rest of the months, all energy consumption data points are low for December, January, and February. A smaller proportion of energy consumption data with higher values from March – November might be due to the use of a refrigerator and washing machine during spring, summer, and fall, which might not be the case in winter. As shown in Figure 10, for sub metering 3, the median energy consumption is near zero from midnight till 06:00 a.m. due to the electric water heater being idle during the nighttime. Energy consumption increases from 07:00 a.m. – 11:00 a.m. due to the use of a water heater. From noon till 11:00 p.m., the median energy decreases. Concerning days of the week, the median energy consumption remains constant. The median energy consumption over the months remains constant. For July and August, a large proportion of energy consumption data points are high and probably identical. However, for the remaining months, a large proportion of energy consumption data points are low and probably identical. This is because of the use of air conditioning during July and August (i.e., summer in France).

3) TRENDS AND SEASONALITY OF THE FEATURES DATA DISTRIBUTION

We decompose each feature of the IHEPC dataset to identify trends and seasonality in the time series data. For each feature, the data is decomposed into 4 components: 1) observed data representing the average values of the data series, 2) trend, indicating an increasing or decreasing behavior of data series over time, 3) seasonality representing repeating cycles or patterns of behavior over time, and 4) residual showing random variation in the time series. The dataset under study contains more than 2 million data points sampled every minute. For better visualization of decomposed components, we resampled the dataset with monthly frequency by adding minutely sampled data for each month. We then performed data decomposition on the resampled dataset.

Figure 11 shows the decomposed data for global active power. Active power consumption has an increasing trend until January 2008, followed by a decrease until July 2008,

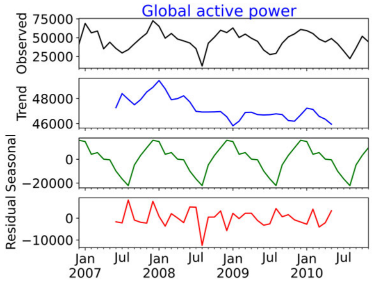

FIGURE 11. Data decomposition of global active power.

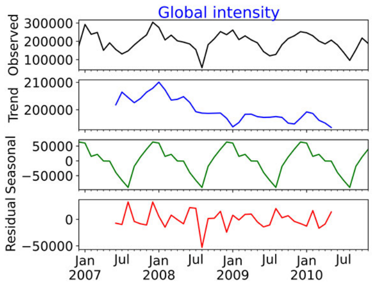

FIGURE 12. Data decomposition of global intensity.

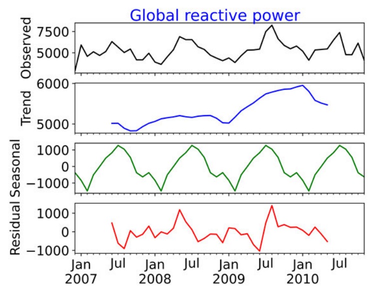

FIGURE 13. Data decomposition of global reactive power.

after which it becomes almost constant. Furthermore, global active power shows a strong component of seasonality. At the beginning of each year, power consumption decreases until August and then increases until December. The decrease in power consumption towards August is due to the weather in France, with spring occurring from March to May, followed by summer from June to August. Consequently, every year during these months the use of an electric water heater is reduced, leading to low power consumption. Similar data trend and seasonality are observed for global intensity (Figure 12) as active power and intensity are highly correlated (Figure 3). Global reactive power, depicted in Figure 13, exhibits an increasing trend with seasonality similar to that of global active power. Figure 14 shows the decomposed components for voltage, with a near-constant trend with strong seasonality.

Figures 15 –17 present the decomposed component for sub metering 1 (i.e., kitchen area), sub metering 2 (i.e., laundry area), and sub metering 3 (i.e., air conditioner and electric

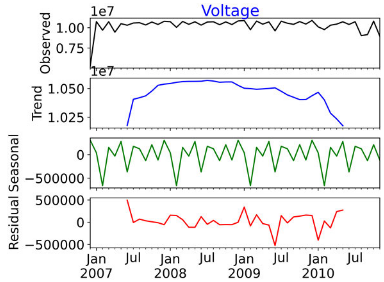

FIGURE 14. Data decomposition of voltage.

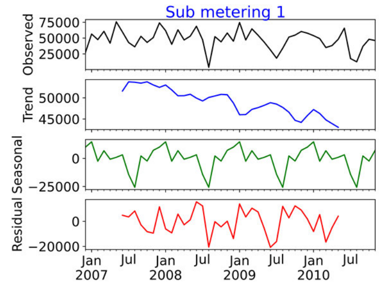

FIGURE 15. Data decomposition of sub metering 1.

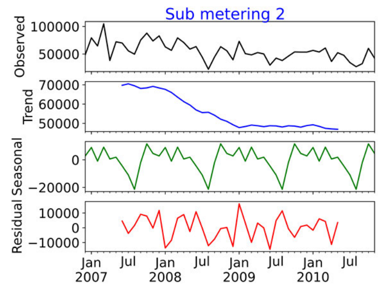

FIGURE 16. Data decomposition of sub metering 2.

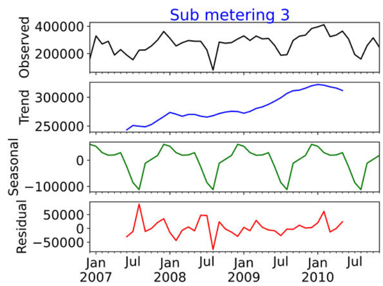

FIGURE 17. Data decomposition of sub metering 3.

water heater) respectively. As shown in Figure 15, the energy consumption of the kitchen area displays a decreasing trend over time, possibly due to the reduced use of dishwashers and microwaves over the years or the adoption of energy-efficient electrical equipment. The consumption, however, demonstrates a strong seasonality. The energy consumption of the laundry area exhibits a decreasing trend until 2008 and

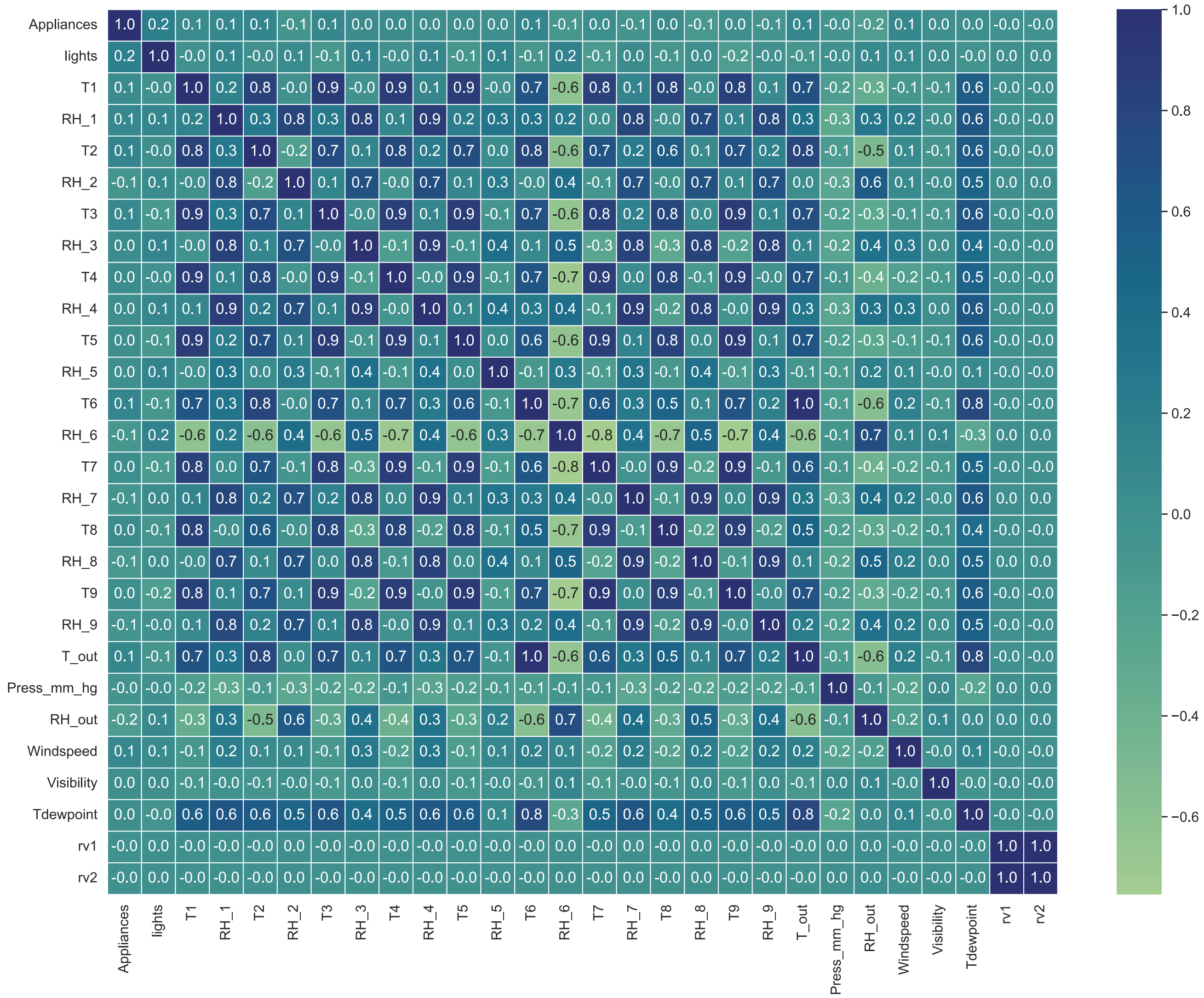

FIGURE 18. **Pearson correlation between features in the AEP dataset.**

then becomes constant for the remaining period as shown in Figure 16. Regarding air conditioning and electric water heater (i.e., sub metering 3), energy consumption shows an increasing trend (Figure 17), possibly due to global warming leading to increased use of air conditioning in summer and electric water heaters in winter. Sub metering 3 shows a strong seasonality with a fixed period, i.e., January – December.

B. AEP DATASET

1) FEATURES CORRELATION WITH APPLIANCES' ENERGY CONSUMPTION

Figure 18 shows the Pearson correlation coefficients [37] between the features in the AEP dataset. As depicted, the temperatures in the kitchen area (T1), living room area (T2), laundry room area (T3), office room (T4), bathroom (T5), outside the building on the north side (T6), ironing room (T7), teenager room 2 (T8), parents' room (T9), outside from the weather station (To), and dew point temperature (Tdewpoint) have significant positive correlations with each other. Temperatures 'T6' and 'To' display a total positive correlation of 1, which is intuitive as both represent the temperature outside the house. Furthermore, the humidities in kitchen area (RH_1), living room area (RH_2), laundry room area (RH_3), office room (RH_4), ironing room (RH_7), teenager room 2 (RH_8), and parents' room (RH_9) show significant positive correlations with each other and weak positive correlations with the humidities in bathroom (RH_5), outside the building on north side (RH_6). Humidity outside the house (RH_out) represents strong positive correlations with 'RH_2', 'RH_6', and 'RH_8', whereas weak positive correlations with 'RH_1', 'RH_3', 'RH_4', 'RH_5', 'RH_7', and 'RH_9'.

Regarding the correlation between humidities and temperatures, the following observations can be made: (i) 'RH_1'

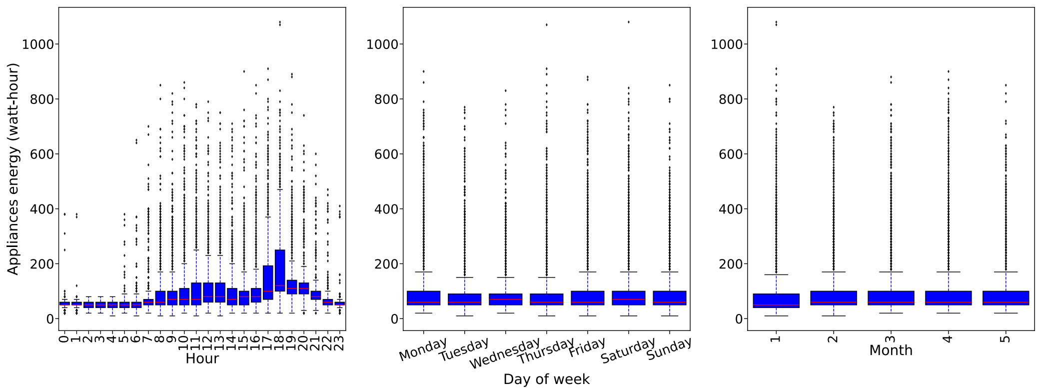

FIGURE 19. Hourly, weekly, and monthly variations in appliances' energy.

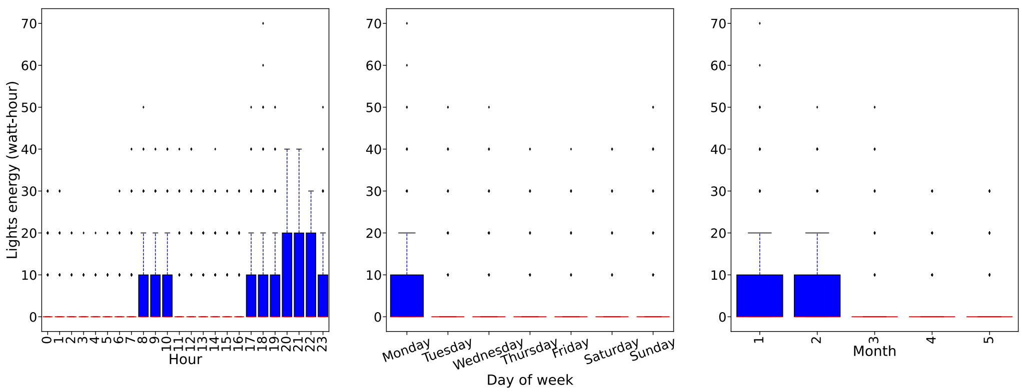

FIGURE 20. Hourly, weekly, and monthly variations in light fixtures energy.

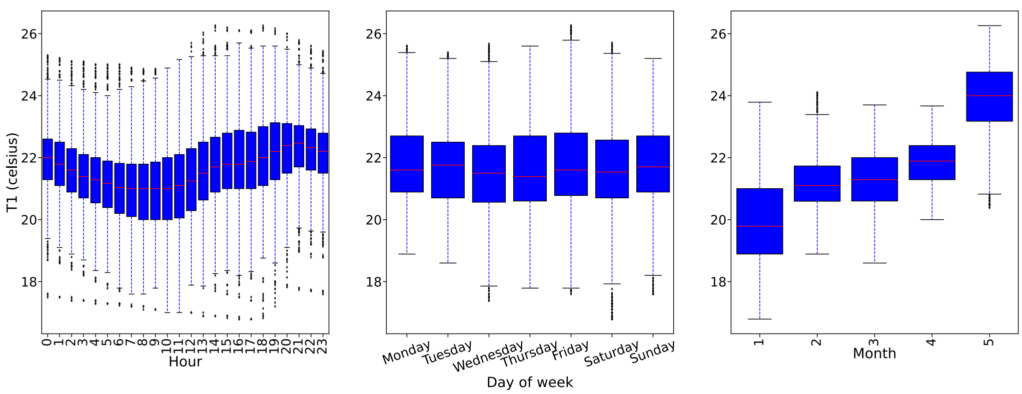

FIGURE 21. Hourly, weekly, and monthly variations in T1 (i.e., kitchen area).

has weak positive correlations with 'T1', 'T2', 'T3', 'T5', 'T6', and 'To', (ii) 'RH_4' and 'RH_9' both have weak positive correlations with 'T1', 'T6', and 'To', (iii) 'RH_7' has weak positive correlations with 'T2', 'T3', 'T6', and 'To'. Moreover, appliances and light fixtures energy consumptions show a weak positive correlation, and random variables 'rv1' and 'rv2' have a correlation of 1. On the other hand, outside humidities (RH_6 and RH_out) demonstrate a strong negative correlation with the temperatures in the kitchen area, living room area, laundry room area, office room, bathroom, north side of the house, teenager room 2, parents' room, and outside the weather station. This is because when the humidity outside the house increases, the temperature decreases due to an increase in moisture content in the air.

2) TEMPORAL DISPERSION OF DATASET FEATURES

Figures 19 – 46 provide insights into the temporal variations of the AEP dataset features for three different time horizons:

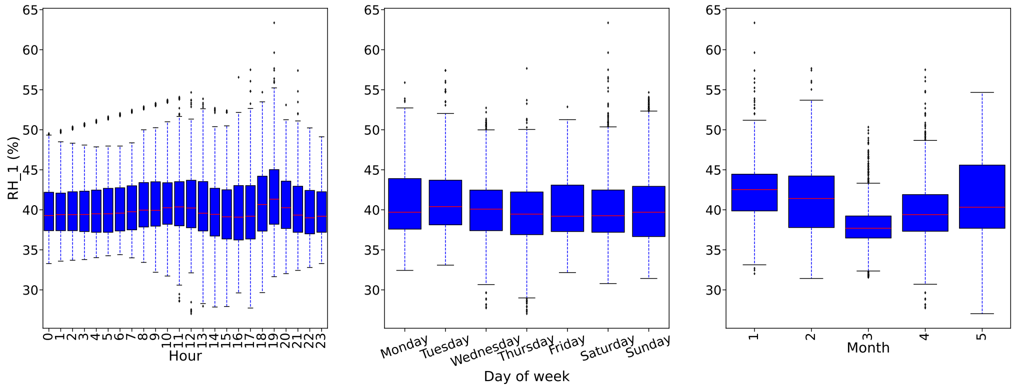

FIGURE 22. Hourly, weekly, and monthly variations in RH_1 (i.e., kitchen area).

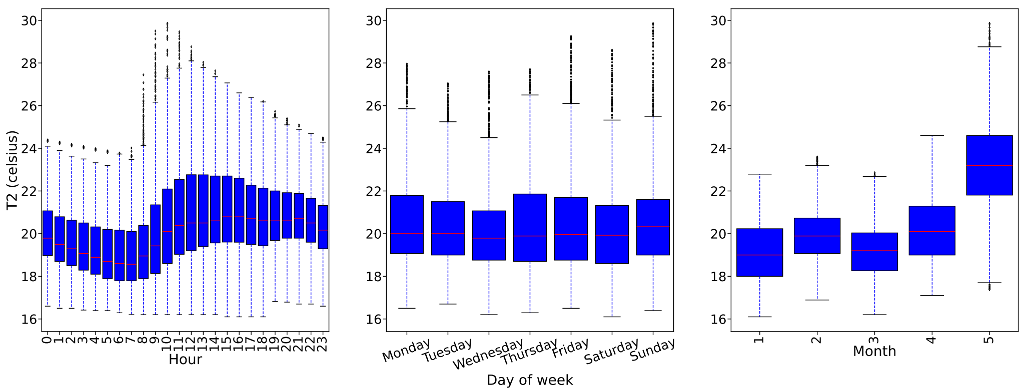

FIGURE 23. Hourly, weekly, and monthly variations in T2 (i.e., living room area).

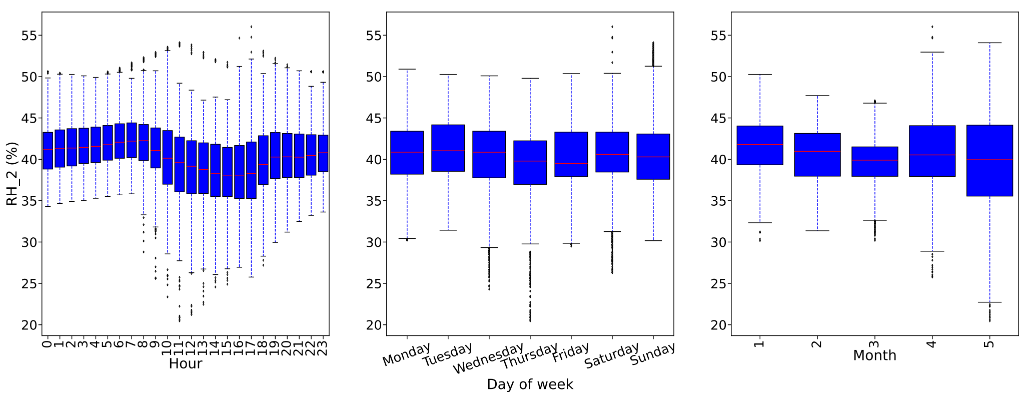

FIGURE 24. Hourly, weekly, and monthly variations in RH_2 (i.e., living room area).

(i) intra-day variations over hours, (ii) weekly variations over days of the week, and (iii) seasonal variations over months. As shown in Figure 19, the appliances' energy consumption from 08:00 p.m. till 07:00 a.m. follows a normal distribution, with nearly equal whisker lengths on both sides of each box. Conversely, from 08:00 a.m. to 07:00 p.m. the distribution is positively skewed. The energy consumption remains constant with the lowest median between 11:00 p.m. and 06:00 a.m., corresponding to the sleeping hours. As the day progresses, energy usage rises from 07:00 a.m. to 01:00 p.m., potentially due to increased appliance use. A slight dip in consumption is observed at 02:00 p.m. during office hours. As people return from work/school, consumption gradually increases after 04:00 p.m. and peaks at 06:00 p.m. (i.e., preparation of dinner), and then decreases till 11:00 p.m. Concerning the days of the week, the median power consumption is relatively constant, with slight increases on Wednesday and Saturday. On a seasonal basis, median power consumption

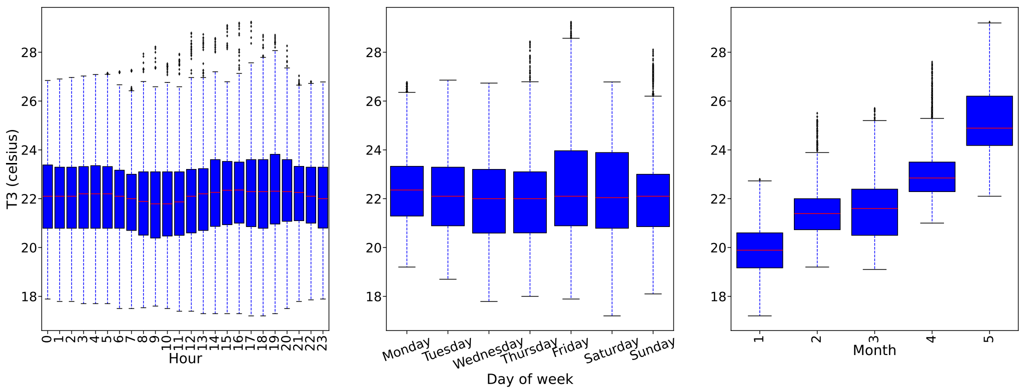

FIGURE 25. Hourly, weekly, and monthly variations in T3 (i.e., laundry room area).

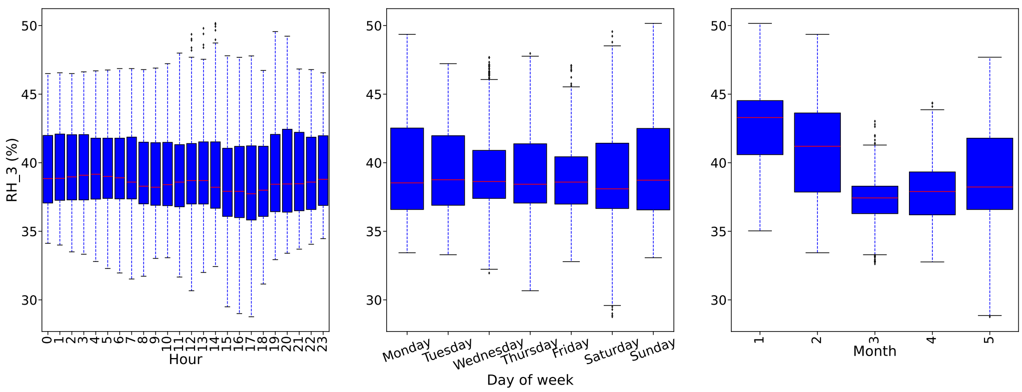

FIGURE 26. Hourly, weekly, and monthly variations in RH_3 (i.e., laundry room area).

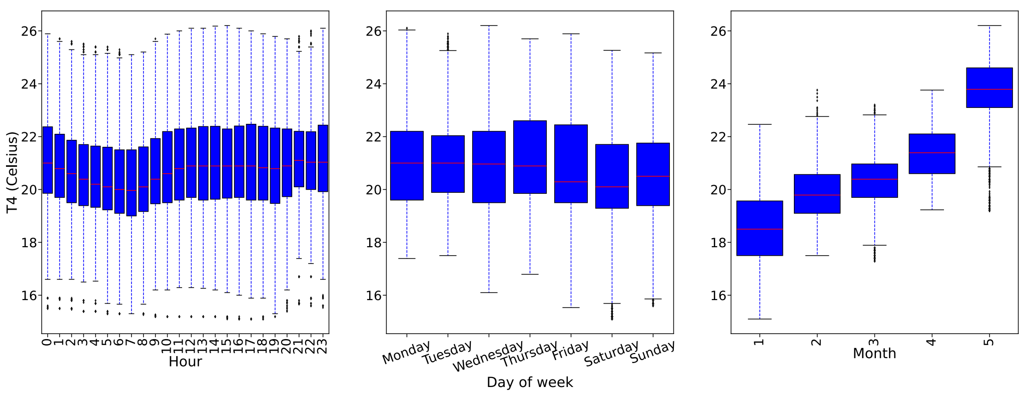

FIGURE 27. Hourly, weekly, and monthly variations in T4 (i.e., office room).

remains constant throughout the data collection period from January to May. The median energy consumption of light fixtures (Figure 20) is almost zero, with increased usage observed in the morning from 8 – 10 a.m. and evening from 5 – 11 p.m. This is because of the use of light fixtures during the morning time and evening time. Energy consumption between 11:00 a.m. and 04:00 p.m. is lower, likely due to reduced occupancy during office/school hours. Throughout the week and over the months, median energy consumption remains stable. However, higher consumption in January and February is observed due to the shorter duration of sunlight during these months in the city of Mons.

The temperature in the kitchen area, depicted in Figure 21, exhibits a normal distribution with the mean temperature decreasing from midnight until 06:00 a.m. This decline is likely attributed to the absence of occupants in the kitchen during these early hours. The temperature then remains relatively constant from 07:00 – 10:00 a.m.,

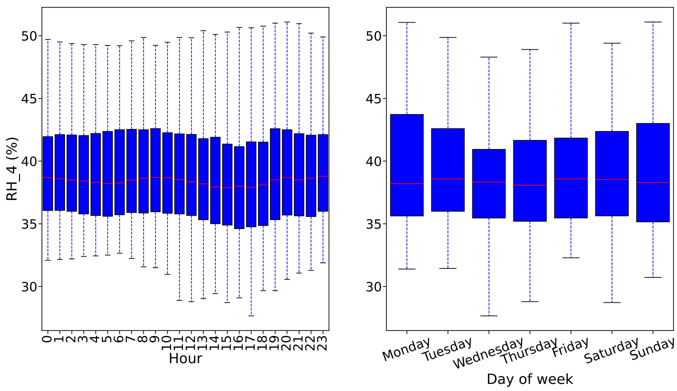

FIGURE 28. Hourly, weekly, and monthly variations in RH_4 (i.e., office room).

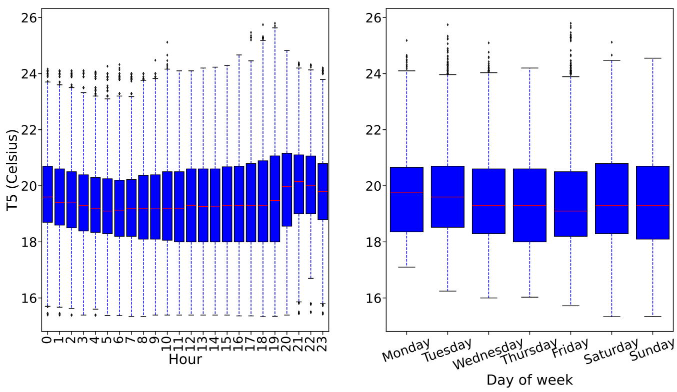

FIGURE 29. Hourly, weekly, and monthly variations in T5 (i.e., bathroom).

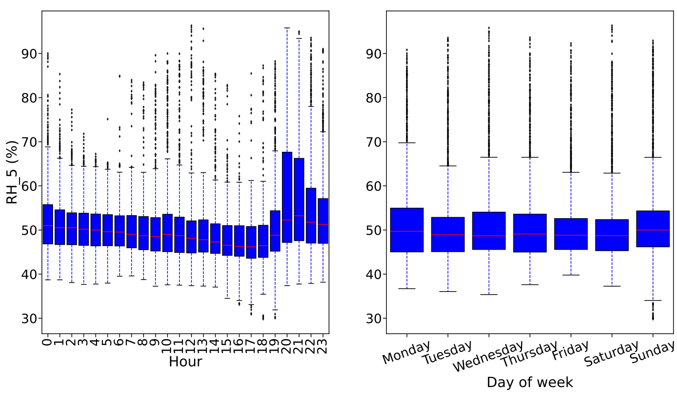

FIGURE 30. Hourly, weekly, and monthly variations in RH_5 (i.e., bathroom).

possibly indicating the influence of a thermostat that regulates the kitchen temperature during breakfast time. The median temperature then increases throughout the day, peaking at 09:00 p.m. and subsequently decreasing. The median temperature remains nearly constant during the week, while an increase is observed from January to May due to the transition from the cool to warm season in Belgium. Figure 22 illustrates the median humidity in the kitchen area, showing a slight increase until 11:00 a.m. This trend may be linked to the inverse relationship between temperature and humidity, given the decreasing kitchen area temperature during this period (Figure 21). From noon the median humidity decreases until 04:00 p.m., followed by an increase until 07:00 p.m. Regarding the variations across the week, the median humidity remains almost constant. Over months, the highest median humidity is observed in January, decreasing until March, and then increasing in April and May.

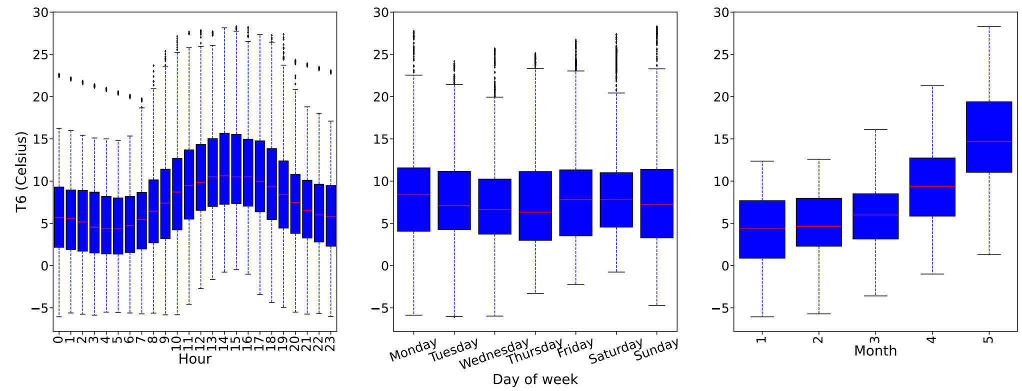

FIGURE 31. Hourly, weekly, and monthly variations in T6 (i.e., north side outside the building).

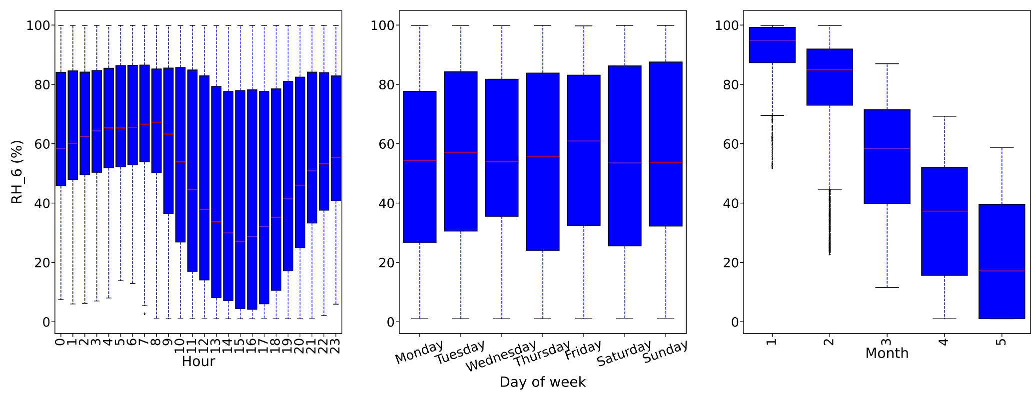

FIGURE 32. Hourly, weekly, and monthly variations in RH_6 (i.e., north side outside the building).

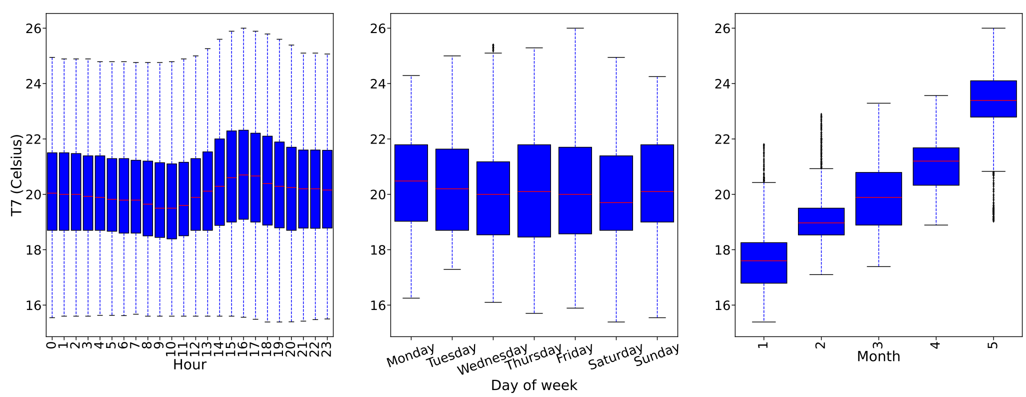

FIGURE 33. Hourly, weekly, and monthly variations in T7 (i.e., ironing room).

Figures 23 and 24 illustrate the temporal variations in the living room area temperature (T2) and humidity (RH_2) respectively. The median temperature, depicted in Figure 23, decreases until 07:00 a.m. and then increases and becomes constant at noon. This can be due to the presence of occupants in the living room during the morning. The temperature further increases after 01:00 p.m. and then remains almost constant with slight fluctuations. The median temperature in the living room decreases after 09:00 p.m. as the occupants might go to their bedrooms from the living room. The weekly analysis reveals a marginal increase in median temperature on Sundays, possibly attributable to weekend activities. Moreover, the temperature displays a rising trend from January to May, aligning with the onset of the warm season in Belgium. In contrast, the humidity in the living room area increases till 08:00 a.m., decreases until 04:00 p.m., and then experiences a subsequent increase (Figure 24). These variations may be influenced by different factors, such as morning activities,

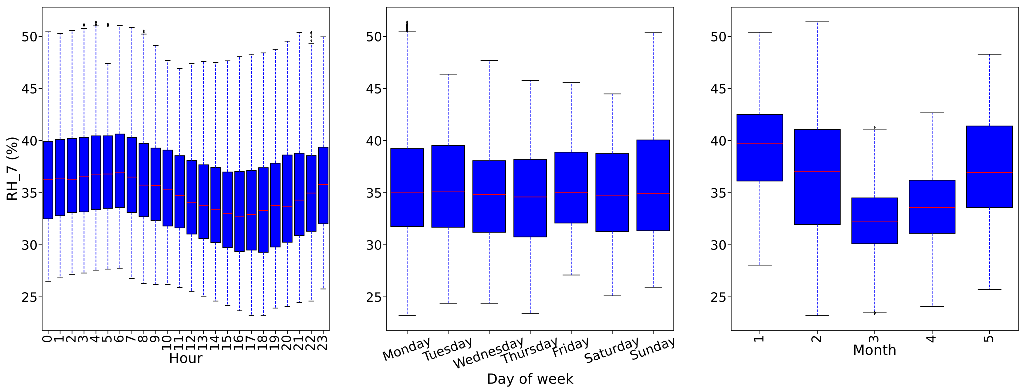

FIGURE 34. Hourly, weekly, and monthly variations in RH_7 (i.e., ironing room).

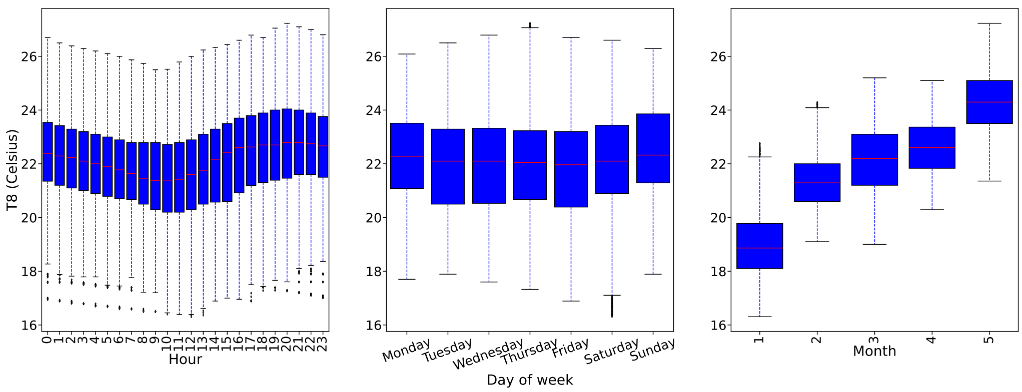

FIGURE 35. Hourly, weekly, and monthly variations in T8 (i.e., teenager room 2).

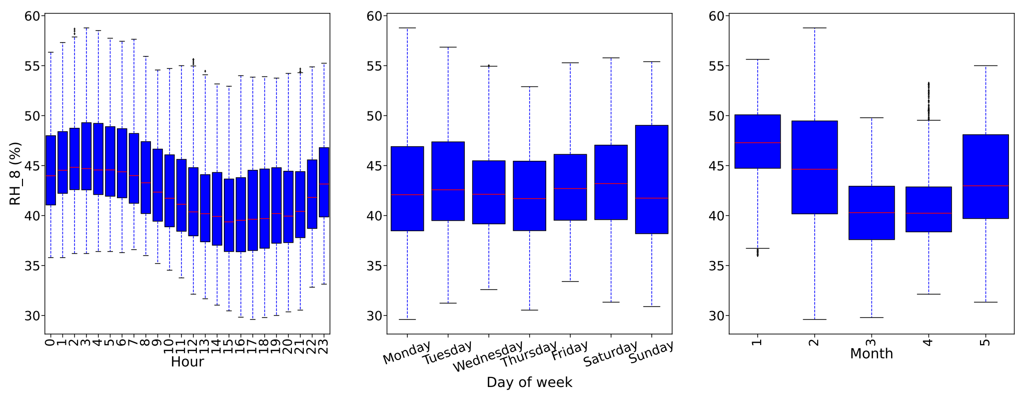

FIGURE 36. Hourly, weekly, and monthly variations in RH_8 (i.e., teenager room 2).

ventilation, and/or external weather conditions. Despite these fluctuations throughout the day, the median humidity remains relatively constant across the week and over months.

Figure 25 presents the temperature variations in the laundry room area. The corresponding humidity variations are presented in Figure 26, providing insights into the inverse relationship between temperature and humidity. As depicted in Figure 25, the temperature variations within a day exhibit a normal distribution with a nearly constant median value, possibly due to limited occupancy during the day. The median temperature remains constant throughout the week, with increased temperature dispersions on Fridays and Saturdays. This may be indicative of heightened laundry room usage during the weekend. Furthermore, the monthly analysis reveals a gradual increase in the temperature from January to May. This shift aligns with seasonal changes, potentially influenced by the warmer weather and increased demand for laundry services in the Spring. In contrast, the variations in the laundry

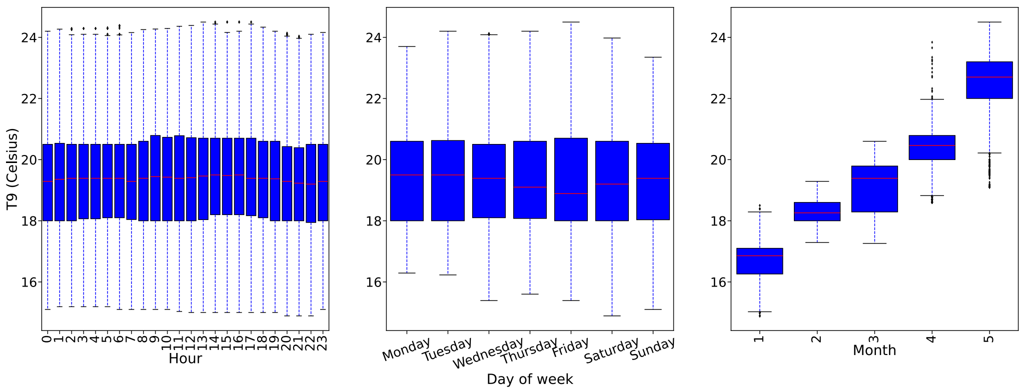

FIGURE 37. Hourly, weekly, and monthly variations in T9 (i.e., parents room).

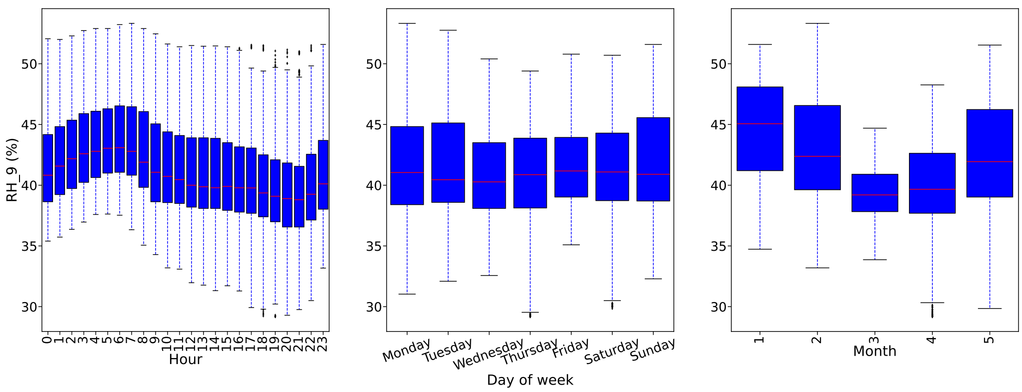

FIGURE 38. Hourly, weekly, and monthly variations in RH_9 (i.e., parents room).

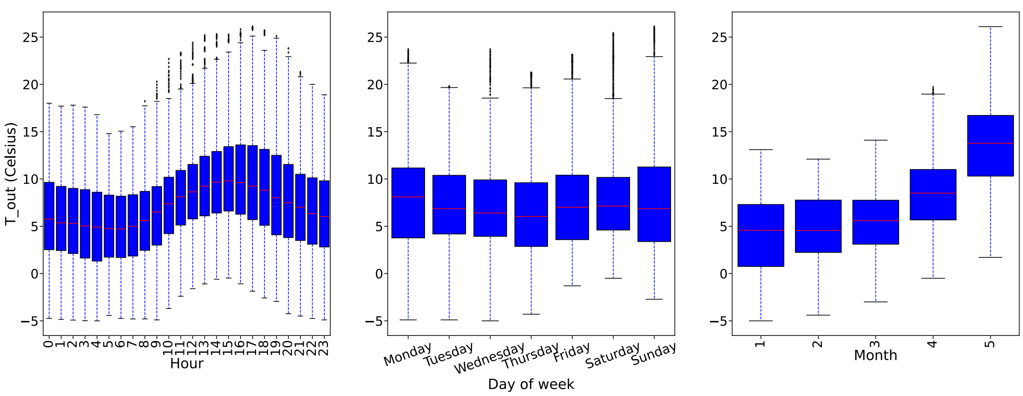

FIGURE 39. Hourly, weekly, and monthly variations in T_out (i.e., outside).

room's humidity (Figure 26) depict opposite trends compared to the temperature (Figure 25). As the temperature decreases, humidity tends to rise, and vice versa. This is due to the inverse relationship between temperature and humidity.

As depicted in Figure 27, the median temperature in the office room decreases from midnight until 07:00 a.m. Later, as work begins, the median temperature increases until noon and then remains constant till 05:00 p.m. After office hours, the median temperature rises, possibly because there might be no one in the office room. The median temperature in the office room remains almost constant from Monday to Thursday. However, on Friday, Saturday, and Sunday, the median temperature is relatively lower compared to other days, reflecting the influence of weekends. The median temperature increases from January to May due to the transition from colder to warmer season in Belgium. On the other hand, relative humidity in the office room (Figure 28) exhibits opposite variations compared to the temperature (Figure 27).

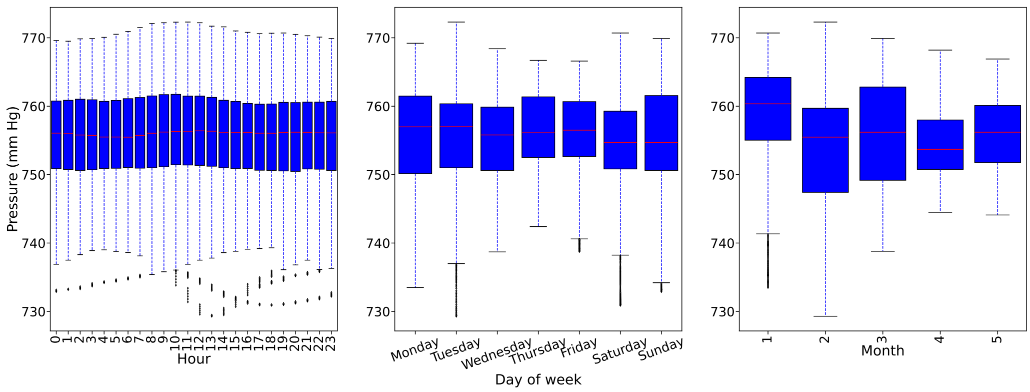

FIGURE 40. Hourly, weekly, and monthly variations in pressure.

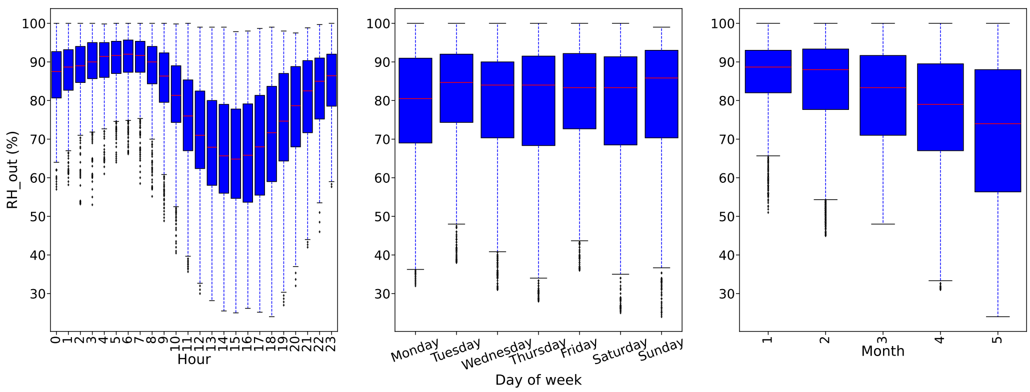

FIGURE 41. Hourly, weekly, and monthly variations in RH_out (i.e., outside).

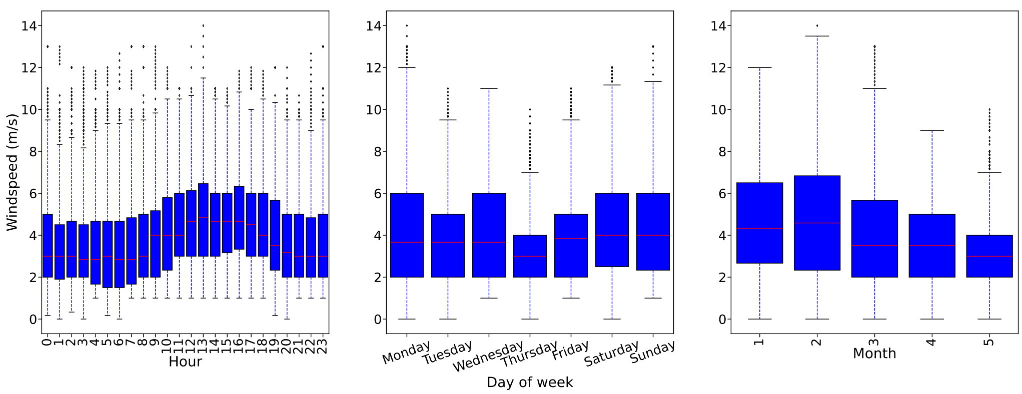

FIGURE 42. Hourly, weekly, and monthly variations in windspeed.

Figure 29 shows the variations in bathroom temperature. It reveals that the temperature follows a normal distribution with a constant median from 08:00 a.m. to 06:00 p.m. The temperature increases in the evening from 07:00 p.m. to 09:00 p.m. This increase might be attributed to the hot water shower taken by the occupants at night before sleeping. The temperature variations remain almost constant during the week but exhibit an upward trend over the months from January to May. Additionally, the relative humidity in the bathroom increases with the rising temperature, as shown in Figure 30. This unusual relationship between temperature and humidity is due to the hot water shower, which contributes to an increase in both temperature and humidity in the bathroom.

The variations in temperature and relative humidity outside the house (north side) depict opposite trends as shown in Figures 31 and 32 respectively. The outside temperature increases during the day and decreases during the night, while the humidity is higher during the night and lower during the

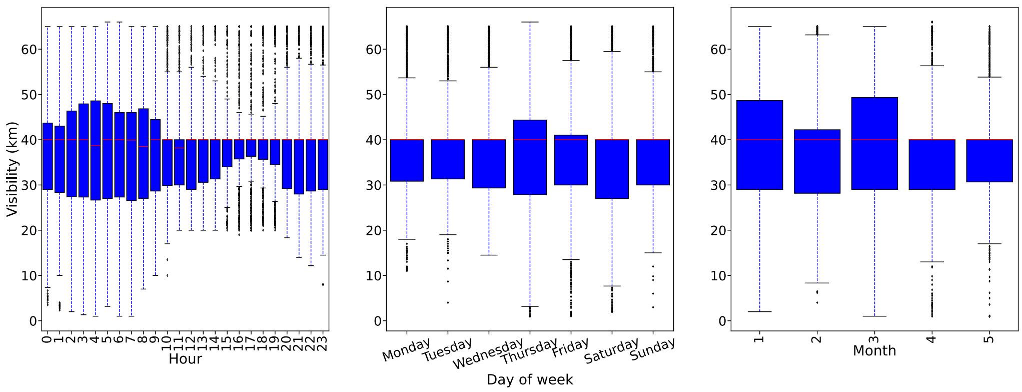

FIGURE 43. Hourly, weekly, and monthly variations in visibility.

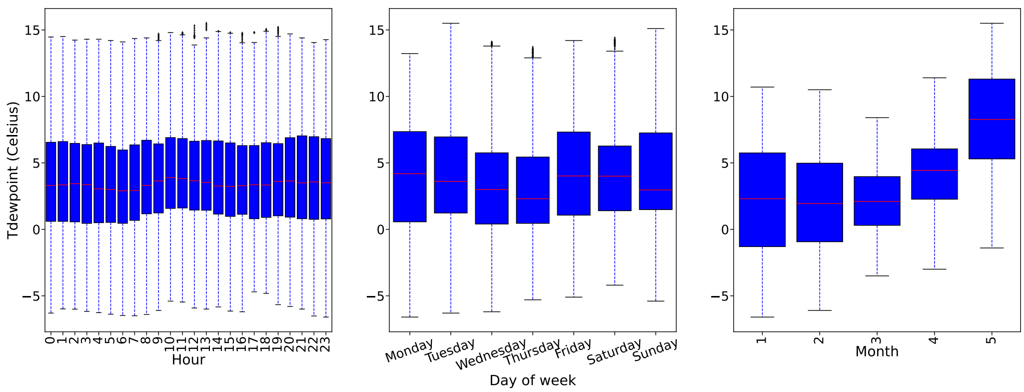

FIGURE 44. Hourly, weekly, and monthly variations in Tdewpoint.

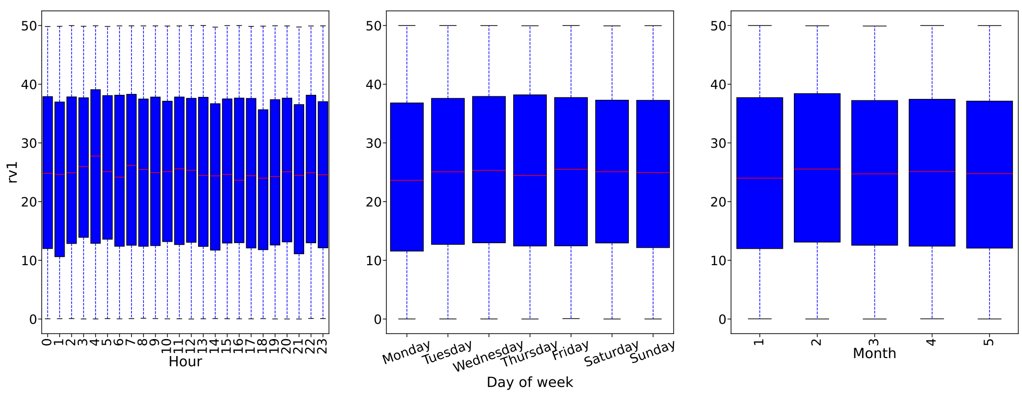

FIGURE 45. Hourly, weekly, and monthly variations in rv1.

day, which is intuitive given the higher temperatures during the day. Both temperature and humidity remain almost constant throughout the week. Over the months, from January to May, the temperature increases due to the arrival of the warm season, and the humidity decreases as the weather becomes dry. Similarly, the temperature and humidity variations in the ironing room present opposing trends (Figures 33 and 34).

The temperature variations in teenager room 2 present a normal distribution, as shown in Figure 35. The temperature decreases during the day, possibly because occupants are attending school or college. The median temperature starts increasing in the afternoon as the occupants return home. Due to the weekend, the median temperature in the teenager room 2 is higher on Sunday compared to other days. The temperature increases over the months due to seasonal changes. In contrast, the relative humidity variations in teenager room 2 show opposite trends (Figure 36). The temperature variations in the parents' room exhibit a normal distribution,

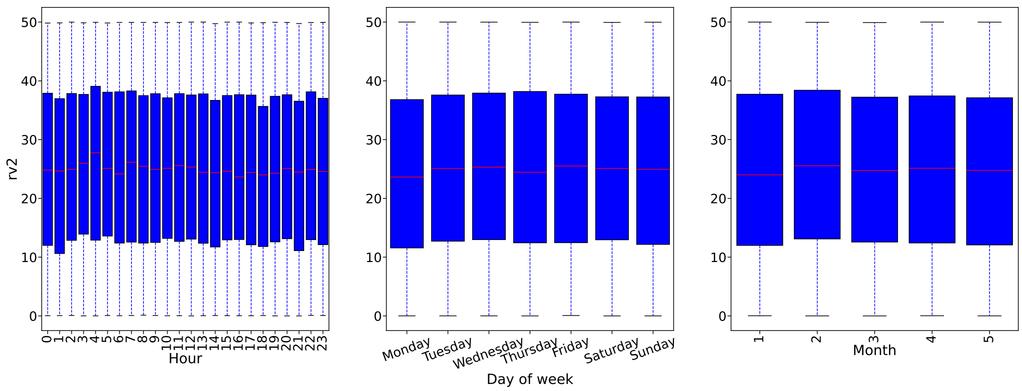

FIGURE 46. Hourly, weekly, and monthly variations in rv2.

as presented in Figure 37. The temperature remains constant throughout the day and across the week. However, due to seasonal changes, the temperature increases from January to May. The relative humidity in the parents' room as depicted in Figure 38, is high during the night and lower during the day. This is because, at night, the wind is generally cooler, leading to an increase in humidity. The variations across the week remain almost constant. The median humidity in the parents' room is the highest in January and the least in April.

The variations in outside temperature follow a normal distribution, with higher temperatures during the day due to the presence of sunlight and lower during the night (Figure 39). The median outside temperature remains almost constant across the week but increases from January to May due to seasonal changes. As shown in Figure 40, the pressure variations also follow a normal distribution, with almost constant median values throughout the day and across the week. However, the median pressure decreases from January to May. This is because in Belgium, being a mid-latitude country, the warmer temperature in May compared to January can contribute to lower pressure. Another reason is that in January, the temperature difference between the cold continental air masses over Europe and the warmer oceanic air masses can create stronger pressure. However, in May, these temperature differences are significantly less, resulting in weaker pressure. The normal distribution of the variations in outside humidity (Figure 41) exhibits opposite trends compared to the outside temperature (Figure 39). Figure 42 represents the variations in windspeed. The median windspeed is higher in the noon and evening compared to the late evening and early morning. It remains almost constant throughout the week. The median windspeed decreases from January to May. This is because windspeed depends on pressure, and the lower pressure in May compared to January (Figure 40) results in lower windspeed (Figure 42).

Figure 43 shows that variations in visibility are negatively skewed, with almost constant median values throughout the day, across the week, and over months. Figure 44 reveals that variations in dewpoint temperature follow a normal distribution with almost constant median values throughout the day.

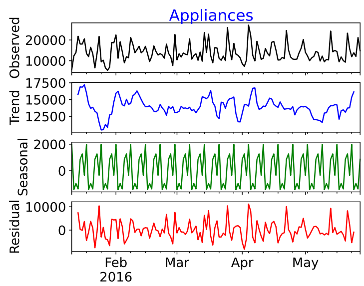

FIGURE 47. Data decomposition of appliances energy.

The median dewpoint temperature increases from January to May as the outside temperature is higher in May. Figures 45 and 46 present the variations for random variables 'rv1' and 'rv2'. As depicted, the random variables exhibit a normal distribution with the median values almost constant throughout the day, across the week, and over months.

3) TRENDS AND SEASONALITY OF THE FEATURES DATA DISTRIBUTION

We decompose each feature of the AEP dataset to identify a trend and seasonality in the time series data. The dataset contains 19735 data points sampled every 10 minutes. For a better visualization of decomposed components, we resampled the dataset with daily frequency by adding minutely sampled data for each day. We then performed data decomposition on the resampled dataset.

Figure 47 shows the decomposed data for appliances' energy consumption. As depicted, energy consumption exhibits a fluctuating trend with the least energy around mid-January. Furthermore, energy consumption displays a strong component of seasonality. At the beginning of each week, energy consumption is at its highest. It then decreases from Monday to Wednesday and then increases again on Thursday and Friday. The energy then decreases on Saturday and rises to the maximum again on Sunday. The energy consumption of light fixtures shows a decreasing trend (Figure 48). This is because of the increase in sunlight presence from January to

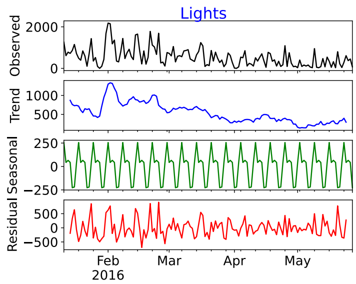

FIGURE 48. **Data decomposition of light fixtures energy.**

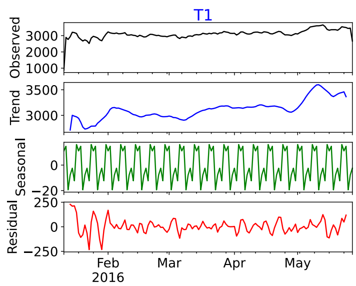

FIGURE 49. **Data decomposition of T1.**

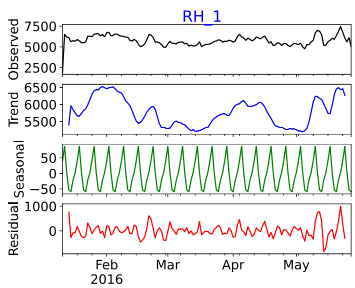

FIGURE 50. **Data decomposition of RH_1.**

May. The light fixtures' energy consumption exhibits weekly seasonality, decreasing during the week and increasing on the weekends. Figure 49 presents the decomposed data for the kitchen area temperature. Due to seasonal changes, the temperature follows an increasing trend from January to May. The temperature also exhibits a strong seasonality, with the highest temperature every weekend. The trend for kitchen area humidity exhibits no trend with seasonality as depicted in Figure 50.

The temperature in the living room area demonstrates an increasing trend, as shown in Figure 51. The temperature has a strong seasonality component, with fluctuating temperatures on weekdays and higher temperatures on the weekends. The humidity in the living room area shows no trend but has seasonality, with decreasing humidity at the beginning of each week and increasing humidity towards the end (Figure 52). Figure 53 presents the decomposition of temperature in the laundry room. As shown, the temperature has an increasing trend with a strong seasonality. In contrast,

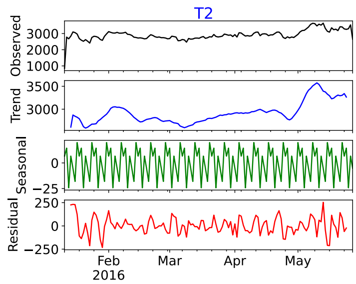

FIGURE 51. **Data decomposition of T2.**

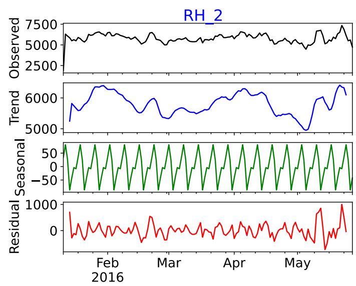

FIGURE 52. **Data decomposition of RH_2.**

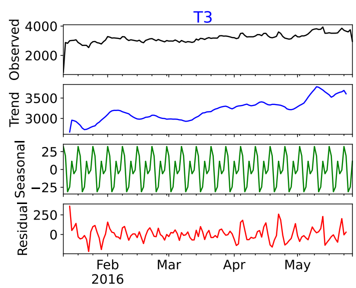

FIGURE 53. **Data decomposition of T3.**

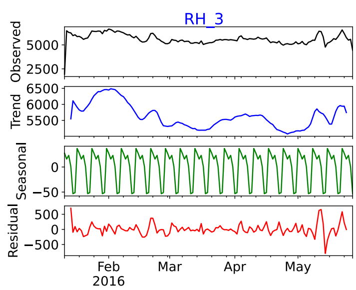

FIGURE 54. **Data decomposition of RH_3.**

humidity in the laundry room exhibits a decreasing trend with seasonality (Figure 54). The humidity decreases during the week and increases as the weekend approaches. Similarly, the temperature and humidity in the office room exhibit an increasing and decreasing trend respectively (Figures 55 and 56). Both temperature and humidity have

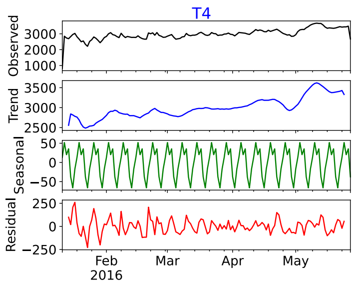

FIGURE 55. Data decomposition of T4.

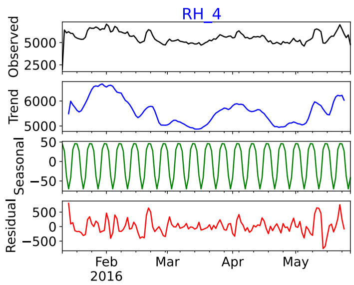

FIGURE 56. Data decomposition of RH_4.

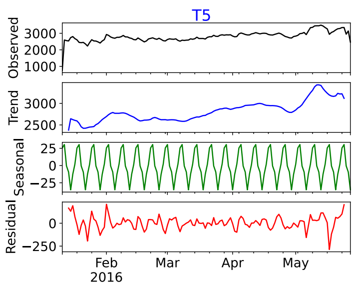

FIGURE 57. Data decomposition of T5.

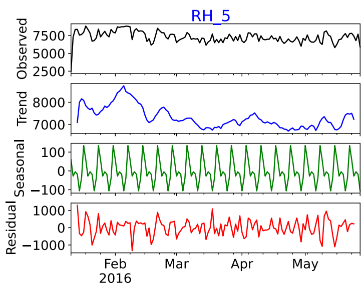

FIGURE 58. Data decomposition of RH_5.

strong seasonality components, with high temperatures every weekend and low humidity.

Figure 57 depicts the decomposed data for bathroom temperature. It shows that the temperature has an increasing trend with seasonality, representing higher temperatures on weekends compared to weekdays. The decomposition for bathroom humidity shows a decreasing trend with seasonality (Figure 58). The temperature and humidity on the

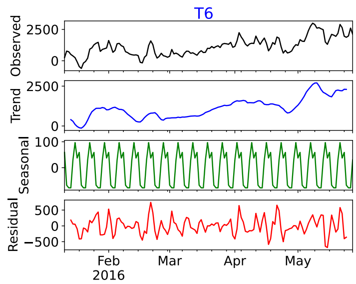

FIGURE 59. Data decomposition of T6.

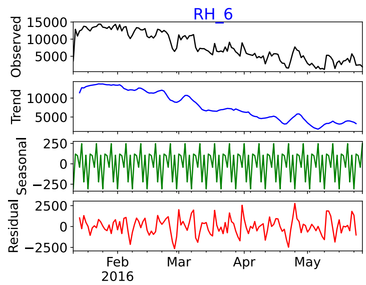

FIGURE 60. Data decomposition of RH_6.

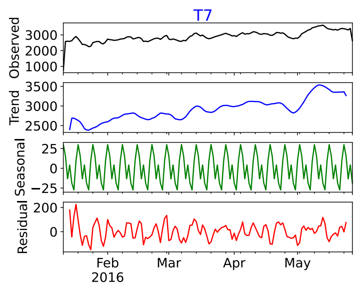

FIGURE 61. Data decomposition of T7.

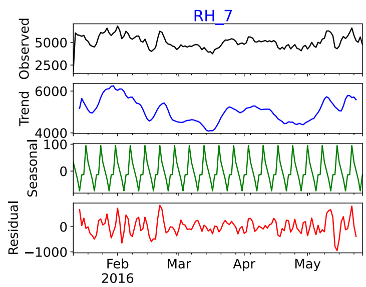

FIGURE 62. Data decomposition of RH_7.

north side of the house exhibit increasing and decreasing trends respectively (Figures 59 and 60). They show a strong seasonality. Similarly, the temperatures in ironing room (Figure 61), teenager room 2 (Figure 63), and parents' room (Figure 65) have increasing trends, whereas the humidities in ironing room (Figure 62), teenager room 2 (Figure 64), and parents' room (Figure 66) have decreasing trends. These

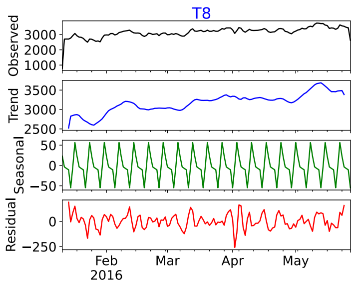

FIGURE 63. Data decomposition of T8.

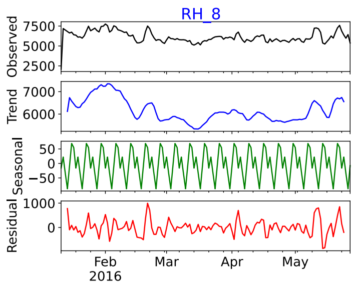

FIGURE 64. Data decomposition of RH_8.

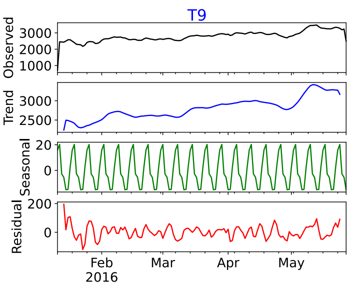

FIGURE 65. Data decomposition of T9.

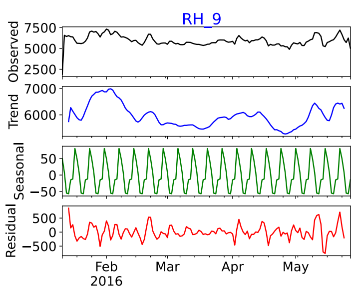

FIGURE 66. Data decomposition of RH_9.

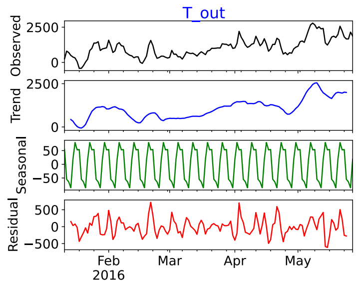

FIGURE 67. Data decomposition of T_out.

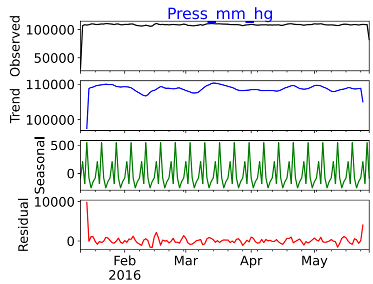

FIGURE 68. Data decomposition of pressure.

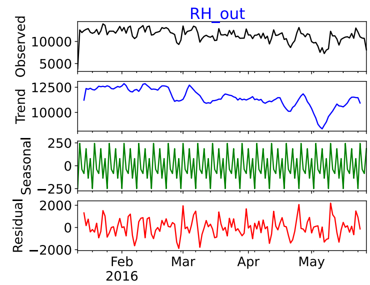

FIGURE 69. Data decomposition of RH_out.

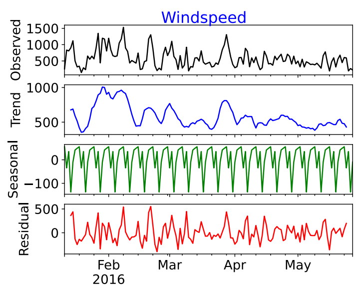

FIGURE 70. Data decomposition of windspeed.

temperatures and humidities demonstrate strong seasonality. Furthermore, the outside temperature and humidity have increasing and decreasing trends, respectively, with seasonality (Figures 67 and 69).

Figure 68 reveals that pressure has a near-constant trend with strong seasonality. In contrast, windspeed (Figure 70) and visibility (Figure 71) demonstrate a decreasing trend with seasonality. Figure 72 represents the decomposed data for dewpoint temperature. As shown, the temperature has an increasing trend with seasonality. The random variables 'rv1' and 'rv2' have fluctuating trends with seasonality as shown in Figures 73 and 74.

V. PERFORMANCE ANALYSIS

In this section, we analyze and compare the performance of LSTM and SVR for forecasting sub metering energy and

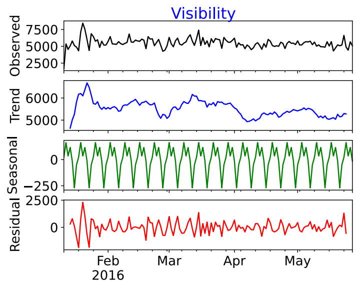

FIGURE 71. Data decomposition of visibility.

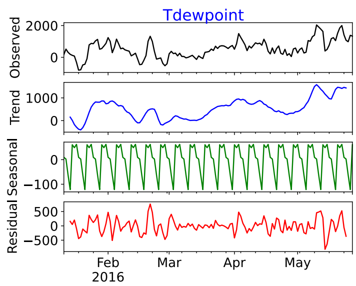

FIGURE 72. Data decomposition of Tdewpoint.

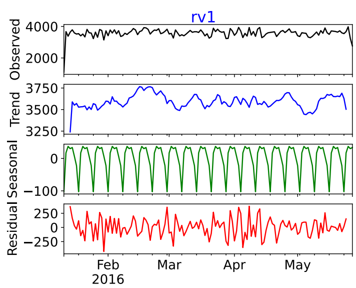

FIGURE 73. Data decomposition of rv1.

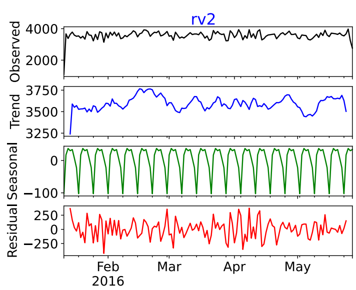

FIGURE 74. Data decomposition of rv2.

global power consumption (using the IHEPC dataset) and appliances' energy consumption (using the AEP dataset) at different forecast levels.

A. EXPERIMENTAL ENVIRONMENT

To evaluate the performance of the most used LSTM and SVR algorithms for different forecast levels, we use the IHPEC and AEP datasets. All the experiments are performed using Python 3.8 on a workstation with AMD Epyc 7552 48-core processor (dual CPU), 1.0 TiB memory, 8.7 TB disk capacity, 2 x NVIDIA RTX A6000 graphics processor with 48 GB memory each, and Ubuntu 22.04.1 LTS operating system.

B. EXPERIMENTS

For each forecast level considered in the IHEPC dataset, i.e., VSTLF (hourly sampling), STLF (daily and weekly samplings), and MTLF (monthly and quarterly samplings), we performed 5 sets of experiments as follows.

1) Sub metering 1: The energy consumption of sub metering 1 (i.e., kitchen area consisting of a dishwasher, an oven, and a microwave) is forecasted using previous sub metering 1 energy consumption values. This is by considering univariate sub metering 1 data from $IHEPC_{hour}$, $IHEPC_{day}$, $IHEPC_{week}$, $IHEPC_{month}$, and $IHEPC_{quarter}$ datasets.
2) Sub metering 2: The energy consumption of sub metering 2 (i.e., laundry area consisting of a washing machine, a tumble-drier, refrigerator, and a light) is forecasted using previous sub metering 2 energy consumption values. This is by considering univariate sub metering 2 data from $IHEPC_{hour}$, $IHEPC_{day}$, $IHEPC_{week}$, $IHEPC_{month}$, and $IHEPC_{quarter}$ datasets.
3) Sub metering 3: The energy consumption of sub metering 3 (i.e., an electric water heater and an air-conditioner) is forecasted using previous sub metering 3 energy consumption values. This is by considering univariate sub metering 3 data from $IHEPC_{hour}$, $IHEPC_{day}$, $IHEPC_{week}$, $IHEPC_{month}$, and $IHEPC_{quarter}$ datasets.
4) Univariate global active power: The global active power consumption is forecasted using previous global active power consumption values. This is by considering univariate global active power consumption data from $IHEPC_{hour}$, $IHEPC_{day}$, $IHEPC_{week}$, $IHEPC_{month}$, and $IHEPC_{quarter}$ datasets.
5) Multivariate global active power: The global active power consumption is forecasted using global reactive power, voltage, global intensity, sub metering 1, sub metering 2, and sub metering 3 values. This is by considering multivariate data from $IHEPC_{hour}$, $IHEPC_{day}$, $IHEPC_{week}$, $IHEPC_{month}$, and $IHEPC_{quarter}$ datasets.

For each forecast level considered in the AEP dataset, i.e., VSTLF (minute and hourly sampling) and STLF (daily and weekly sampling), we performed the following experiments.

- Multivariate appliances' energy: The appliances' energy consumption is forecasted using light fixtures' energy, temperatures (T1 – T9, To, and Tdewpoint), humidities (RH_1 – RH_9 and RH_out), pressure, windspeed, visibility, and random variables (rv1 and rv2). This is by considering the multivariate data from AEP, AEP_{hour}, AEP_{day}, and AEP_{week} datasets.

TABLE 6. Value(s) of hyperparameters used in literature and our experiments for the algorithms under study.

Algorithm	Parameters	Value(s) used in the literature	Values used in our experiments
Long Short-Term Memory (LSTM)	Epochs	Not reported [7], [20]; 16 [24]; 30 [22], [25]; 35 [12]; 50 [6]	16, 30, 35, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000
	Batch size	Not reported [6], [7], [20]; 1 [22]; 30 [25]; 64 [12], [24]	1, 30, 64, length of training data
	Optimizer	Not reported [6], [20]; RMSprop [24]; Adam [7], [12], [22], [25]	RMSprop, Adam
	Learning rate	Not reported [6], [20], [22]; 0.001 [24]; 1e-4, 5e-4, 1e-3*, 5e-3, 1e-2, 5e-2 [7]; adaptive (initial 0.005) [12], [25]	1e-3
	Layers	Not reported [25]; 2 [7], [12]; 3 [20], [24]; 1, 2, 3, 4 [6]; 7 [22]	1
	Units	Not reported [25]; 64, 16 [24]; 50 [22]; 80 per layer [12]; 2, 4, 8, 16, 32, 64, 128 [6]; 5, 10, 20, 30, 32, 40, 50 [7]; 256, 128, 50 [20]	2, 4, 8, 16, 32, 64, 100, 128, 150, 250, 256 per layer
	Drop out	Not reported [6], [7], [20], [22], [24], [25]; 0.3 [12]	0.1, 0.3
	Activation	Not reported [7], [20], [25]; Tanh [6], [24]; Relu [22]; Linear [12]	Tanh, Relu, Linear, Sigmoid
	Loss function	Not reported [6], [7], [20], [25]; MAE [22]; MSE [12], [24]	MSE
	Epsilon	Not reported [6], [7], [12], [20], [22], [25]; 1e-07 [24]	1e-07
Support Vector Regression (SVR)	Kernel	Not reported [13]; RBF [21], [23], [24]	RBF
	Kernel coefficient (gamma)	Not reported [24]; 0.25 [23]; 0.1, 0.35, 0.4* [21]; 10^{-3}, 10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2}, 10^{3} [13]	10^{-3}, 10^{-2}, 10^{-1}, 0.25, 0.4, 1, 10^{1}, 10^{2}, 10^{3}
	Degree†	Not reported [13]; Not applicable [21], [23], [24]	Not applicable
	Independent term in kernel‡	Not reported [13]; Not applicable [21], [23], [24]	Not applicable
	Tolerance for stopping criteria	Not reported [13], [21], [23]; 0.001 [24]	0.001
	Regularization parameter (c)	Not reported [24]; 1 [23]; 1, 3, 5, 8, 10, 12* [21]; 10^{-3}, 10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2}, 10^{3}, 10^{4}, 10^{5}, 10^{6}, 10^{7}, 10^{8}, 10^{9}, 10^{10}, 10^{11}, 10^{12} [13]	10^{-3}, 10^{-2}, 10^{-1}, 1, 10^{1}, 12, 10^{2}, 10^{3}, 10^{4}, 10^{5}, 10^{6}, 10^{7}, 10^{8}, 10^{9}, 10^{10}, 10^{11}, 10^{12}
	Epsilon	Not reported [13], [21]; 0.1 [23], [24]	0.1
	Shrinking	Not reported [13], [21], [23], [24]	True (default)
	Cache size	Not reported [13], [21]; 100 [23]; 200 [24]	1000
	Maximum iterations within solver	Not reported [13], [21], [23], [24]	-1 (default)

*Optimal parameter value; RBF – Radial Basis Function; †Used for polynomial kernel; ‡Used for polynomial and sigmoid kernels

For each set of experiments and each forecast level, we obtained the optimal parameters for LSTM and SVR algorithms by performing a grid search over different values of hyperparameters. Table 6 presents the hyperparameters for LSTM and SVR, and their corresponding ranges used in literature and our experiments. Tables 7 and 8 show the optimal values of hyperparameters obtained in our experiments for LSTM and SVR for IHEPC and AEP datasets respectively. For each set of experiments, we divided $IHEPC_{hour}$, $IHEPC_{day}$, $IHEPC_{week}$, $IHEPC_{month}$, $IHEPC_{quarter}$, AEP, AEP_{hour}, AEP_{day}, and AEP_{week} datasets each into 70% for training the LSTM and SVR models (i.e., model development) and 30% for validating the developed models (i.e., model validation). We evaluate the forecast performance of the developed models in terms of sMAPE, RMSE, and Jensen-Shannon divergence. The sMAPE and RMSE values are calculated using Equations (1) and (2) respectively. Jensen-Shannon divergence measures the similarity between two probability distributions by computing the average of Kullback-Leibler divergences between the distribution of actual energy values and the distribution of predicted energy values, and between the distribution of predicted values and the distribution of actual values [38].

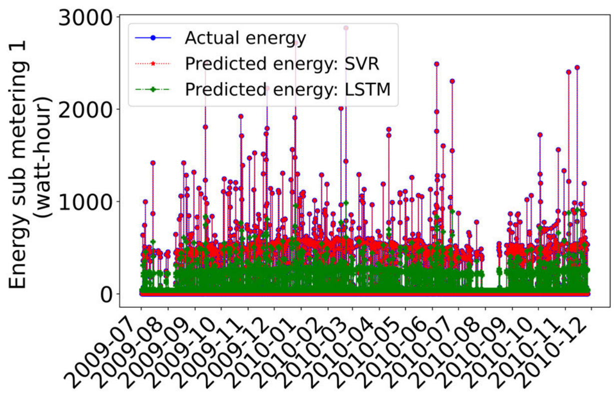

FIGURE 75. Sub metering 1 energy consumption forecast using SVR and LSTM for $IHEPC_{hour}$ dataset.

$$sMAPE = \left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{Actual\ load_t - Forecasted\ load_t}{(|Actual\ load_t| + |Forecasted\ load_t|)/2}\right|\right)\times 100 \quad (1)$$

TABLE 7. Optimal values of hyperparameters obtained in our experiments for the IHEPC dataset.

Algorithm	Sampling	Parameters	Optimal values				
			Sub meter 1	Sub meter 2	Sub meter 3	Global active power (univariate)	Global active power (multivariate)
Long Short-Term Memory (LSTM)	Hour	Epochs	50	100	50	500	500
		Batch size	Length of training data	Length of training data	200	Length of training data	200
		Optimizer	Adam	Adam	Adam	Adam	Adam
		Units	128	32	32	128	32
		Drop out	0.3	0.1	0.1	0.1	0.1
		Activation	relu	Sigmoid	Sigmoid	relu	relu
	Day	Epochs	200	200	200	500	400
		Batch size	120	120	120	Length of training data	100
		Optimizer	Adam	Adam	Adam	Adam	Adam
		Units	128	250	32	32	128
		Drop out	0.1	0.1	0.1	0.1	0.1
		Activation	Sigmoid	Sigmoid	Sigmoid	Sigmoid	relu
	Week	Epochs	500	400	100	500	400
		Batch size	30	10	2	10	10
		Optimizer	Adam	Adam	Adam	Adam	Adam
		Units	32	150	32	32	32
		Drop out	0.1	0.1	0.1	0.1	0.1
		Activation	Sigmoid	Sigmoid	Sigmoid	Sigmoid	relu
	Month	Epochs	100	200	400	500	1000
		Batch size	12	1	1	6	10
		Optimizer	Adam	Adam	Adam	Adam	Adam
		Units	256	100	64	32	256
		Drop out	0.1	0.1	0.1	0.1	0.1
		Activation	Sigmoid	Sigmoid	Sigmoid	Sigmoid	Sigmoid
	Quarter	Epochs	200	500	400	500	600
		Batch size	5	5	1	Length of training data	5
		Optimizer	Adam	RMSprop	Adam	Adam	Adam
		Units	16	2	256	128	128
		Drop out	0.1	0.1	0.1	0.1	0.1
		Activation	Sigmoid	relu	sigmoid	sigmoid	linear
Support Vector Regression (SVR)	Hour	Gamma	1000	1000	1000	100	1000
		C	10000	10000	10000	10000	1000
	Day	Gamma	100	100	100	1	1000
		C	10000	100000	10000	10000	1000
	Week	Gamma	1	1	0.25	0.25	10
		C	100000	100000	100000	10000	10000
	Month	Gamma	0.01	0.1	0.01	0.1	0.001
		C	100000	100000	1000000	100000	100000
	Quarter	Gamma	0.01	0.001	0.01	0.01	0.1
		C	1000000	1000000	1000000	100000	100000

Note – Parameters for which a single value is used in hyperparameter tuning are not mentioned in this table as the used value is the optimal value

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (Actual\ load_t - Forecasted\ load_t)^2}{n}} \tag{2}$$

where n is the total number of records in the validation dataset.

C. EXPERIMENTAL RESULTS ANALYSIS

Figures 75 – 79 present the sub metering 1 energy prediction using SVR and LSTM for $IHEPC_{hour}$, $IHEPC_{day}$, $IHEPC_{week}$, $IHEPC_{month}$, and $IHEPC_{quarter}$ validation datasets respectively. As depicted in the figures, overall SVR outperforms LSTM by predicting the energy consumption

TABLE 8. Optimal values of hyperparameters obtained in our experiments for the AEP dataset.

Algorithm	Sampling	Parameters	Optimal values
Long Short-Term Memory (LSTM)	Min	Epochs	400
		Batch size	Length of training data
		Optimizer	Adam
		Units	32
		Drop out	0.1
		Activation	Linear
	Hour	Epochs	500
		Batch size	150
		Optimizer	Adam
		Units	32
		Drop out	0.1
		Activation	Linear
	Day	Epochs	500
		Batch size	10
		Optimizer	Adam
		Units	32
		Drop out	0.1
		Activation	relu
	Week	Epochs	1000
		Batch size	1
		Optimizer	Adam
		Units	32
		Drop out	0.1
		Activation	Sigmoid
Support Vector Regression (SVR)	Min	Gamma	0.001
		C	10000
	Hour	Gamma	0.1
		C	10000
	Day	Gamma	1
		C	100000
	Week	Gamma	0.001
		C	100000

Note – Parameters for which a single value is used in hyperparameter tuning are not mentioned in this table as the used value is the optimal value

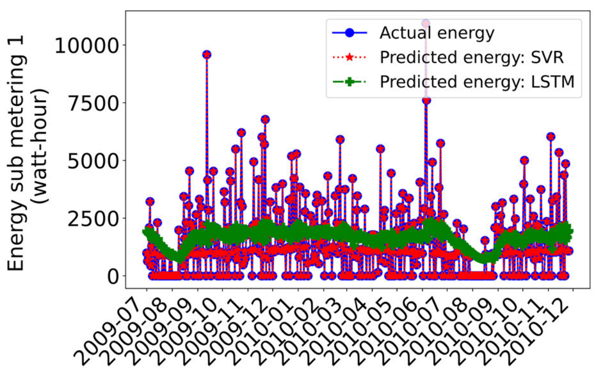

FIGURE 76. Sub metering 1 energy consumption forecast using SVR and LSTM for $IHEPC_{day}$ dataset.

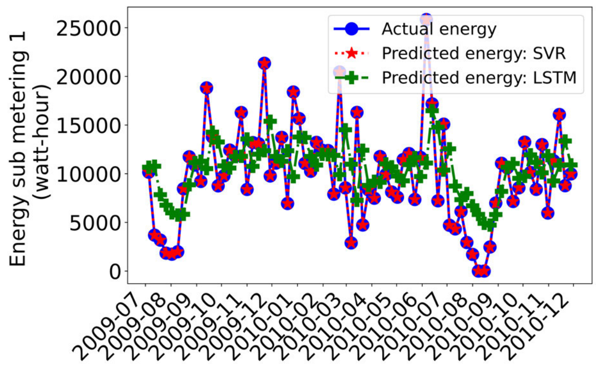

FIGURE 77. Sub metering 1 energy consumption forecast using SVR and LSTM for $IHEPC_{week}$ dataset.

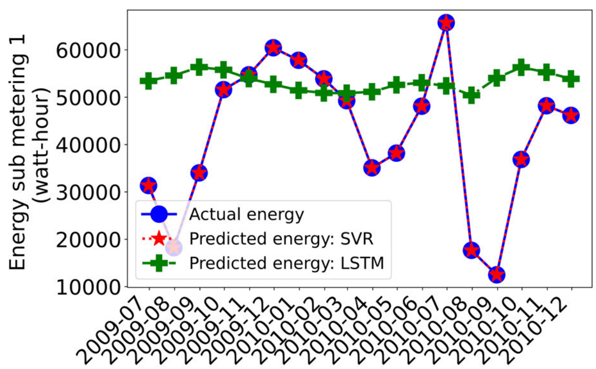

FIGURE 78. Sub metering 1 energy consumption forecast using SVR and LSTM for $IHEPC_{month}$ dataset.

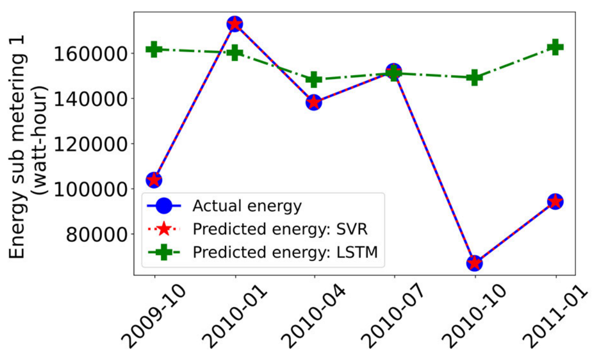

FIGURE 79. Sub metering 1 energy consumption forecast using SVR and LSTM for $IHEPC_{quarter}$ dataset.

accurately. This is because sub metering 1 is a non-linear time series (Figure 15). SVR when used with RBF kernel captures this non-linearity while generating the best-fit regression line, leading to accurate predictions. Comparing the performance of LSTM for different forecast levels (Figures 75 – 79), the model has the best prediction result for the weekly forecast level (Figure 77). This is because compared to $IHEPC_{week}$, most of the values in the training dataset for $IHEPC_{hour}$ is 0. Consequently, most of the predictions by LSTM for hourly forecast level are near 0 as shown in Figure 75. Compared to $IHEPC_{week}$ dataset from which we use 133 records for training, we use 30 and 11 records for training from $IHEPC_{month}$ and $IHEPC_{quarter}$ datasets respectively. Our results reveal that these training dataset sizes are not sufficient enough to train an LSTM model, resulting in poor LSTM predictions for $IHEPC_{quarter}$ (Figure 79) and $IHEPC_{month}$ (Figure 78), compared to $IHEPC_{week}$ (Figure 77). On the other hand, SVR predicts the energy consumption accurately for all the dataset

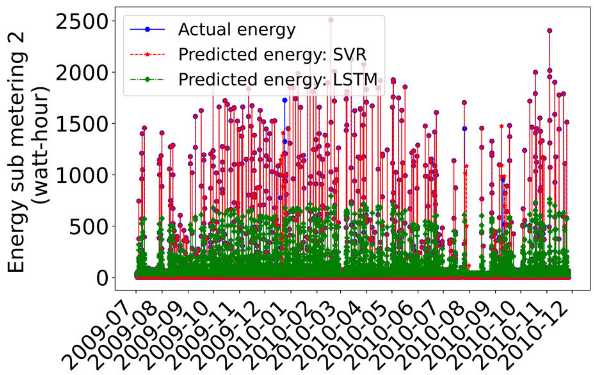

FIGURE 80. Sub metering 2 energy consumption forecast using SVR and LSTM for $IHEPC_{hour}$ dataset.

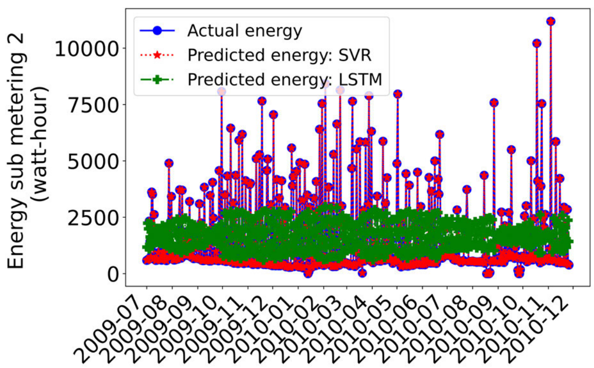

FIGURE 81. Sub metering 2 energy consumption forecast using SVR and LSTM for $IHEPC_{day}$ dataset.

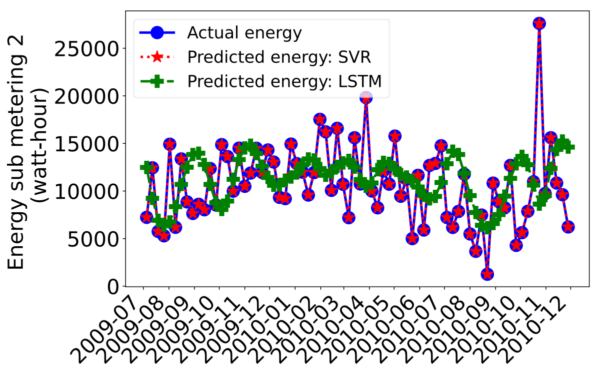

FIGURE 82. Sub metering 2 energy consumption forecast using SVR and LSTM for $IHEPC_{week}$ dataset.

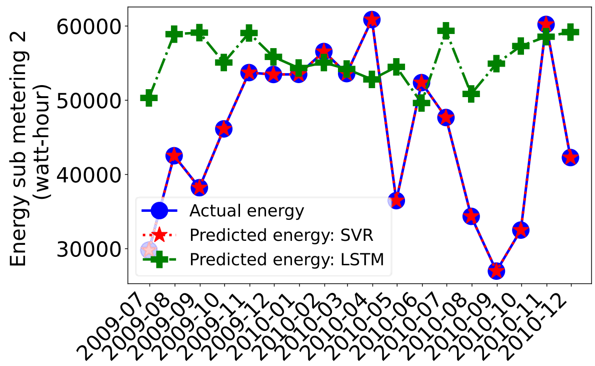

FIGURE 83. Sub metering 2 energy consumption forecast using SVR and LSTM for $IHEPC_{month}$ dataset.

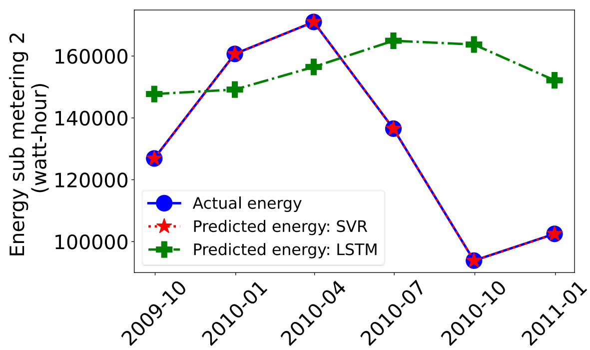

FIGURE 84. Sub metering 2 energy consumption forecast using SVR and LSTM for $IHEPC_{quarter}$ dataset.

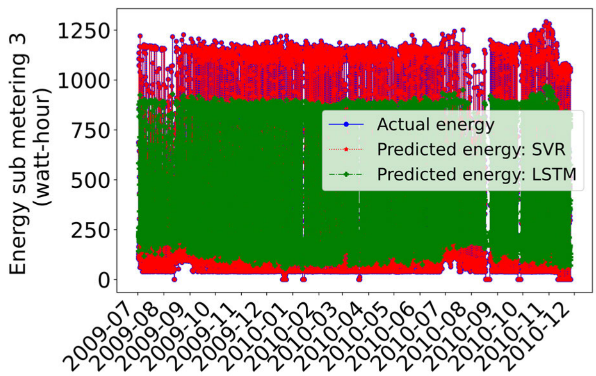

FIGURE 85. Sub metering 3 energy consumption forecast using SVR and LSTM for $IHEPC_{hour}$ dataset.

sizes under study. In summary, LSTM requires larger training dataset for accurate predictions.

Figures 80 – 84 present the sub metering 2 energy prediction using SVR and LSTM for $IHEPC_{hour}$, $IHEPC_{day}$, $IHEPC_{week}$, $IHEPC_{month}$, and $IHEPC_{quarter}$ validation datasets respectively. For all forecast levels, i.e., hourly, daily, weekly, monthly, and quarterly, SVR outperforms LSTM. This is because sub metering 2 time series data trend is non-linear (Figure 16) which is accurately modeled by SVR. Comparing the performances of LSTM for different forecast levels, LSTM has the best performance for hourly (Figure 80), daily (Figure 81), and weekly (Figure 82) forecasts due to a sufficient training dataset size. Poor prediction performance of LSTM for monthly (Figure 83) and quarterly (Figure 84) levels is due to a smaller training dataset size.

Figures 85 – 89 present the sub metering 3 energy prediction using SVR and LSTM for $IHEPC_{hour}$, $IHEPC_{day}$, $IHEPC_{week}$, $IHEPC_{month}$, and $IHEPC_{quarter}$ validation

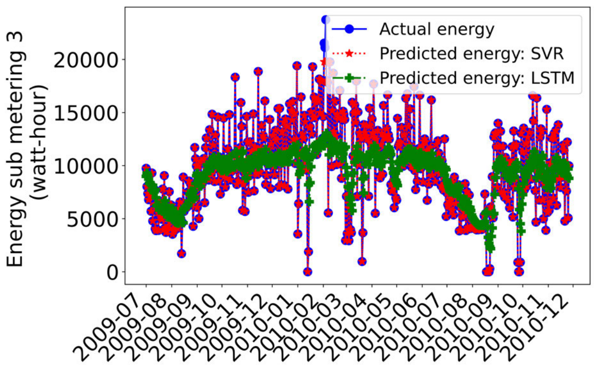

FIGURE 86. Sub metering 3 energy consumption forecast using SVR and LSTM for $IHEPC_{day}$ dataset.

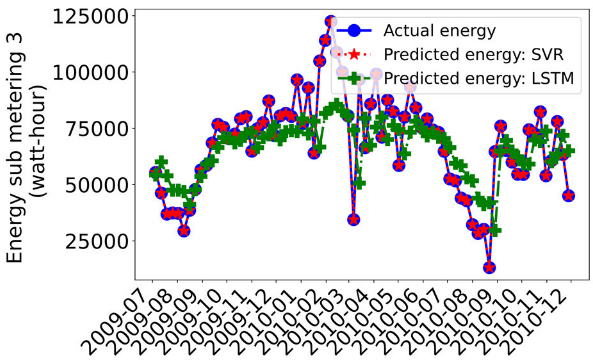

FIGURE 87. Sub metering 3 energy consumption forecast using SVR and LSTM for $IHEPC_{week}$ dataset.

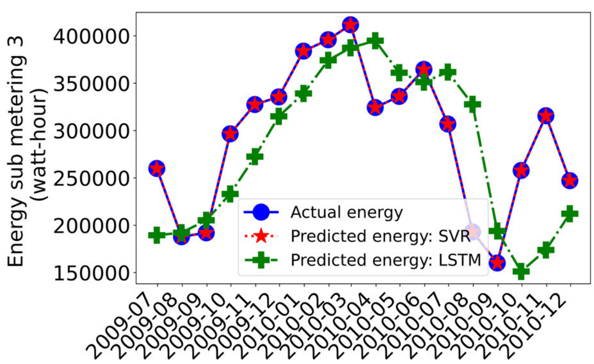

FIGURE 88. Sub metering 3 energy consumption forecast using SVR and LSTM for $IHEPC_{month}$ dataset.

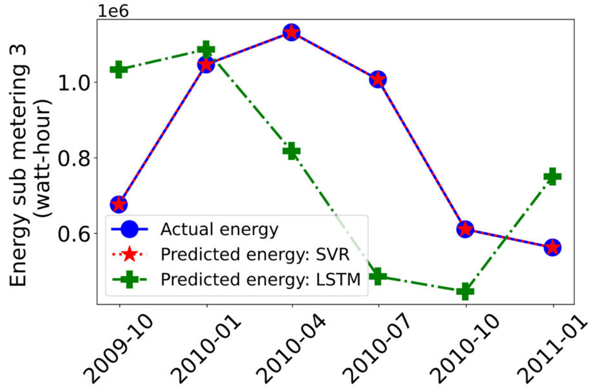

FIGURE 89. Sub metering 3 energy consumption forecast using SVR and LSTM for $IHEPC_{quarter}$ dataset.

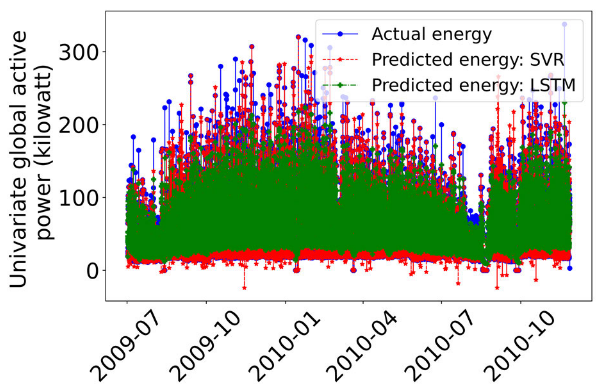

FIGURE 90. Univariate global active power consumption forecast using SVR and LSTM for $IHEPC_{hour}$ dataset.

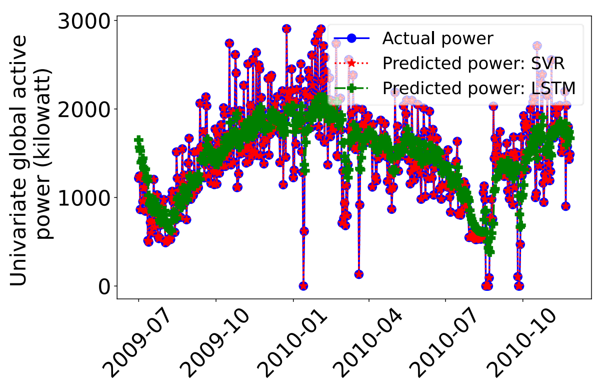

FIGURE 91. Univariate global active power consumption forecast using SVR and LSTM for $IHEPC_{day}$ dataset.

datasets respectively. Figures show that SVR has the most accurate predictions compared to LSTM for all forecast levels. This is because SVR can model the non-linearity that prevails in the data trend for sub metering 3 (Figure 17). Furthermore, comparing the performances of LSTM for different forecast levels, quarterly (Figure 89) depicts high error as the model is trained using small training datasets. Similarly, for univariate global active power consumption, SVR outperforms LSTM (Figures 90 – 94). For each forecast level; i.e., hourly, daily, weekly, monthly, and quarterly, the prediction performance of SVR for sub metering 3 is almost equal to that for univariate global active power as shown in Figures 85 – 89 and 90 – 94 respectively. Similarly, for each forecast level, the prediction performance of LSTM for sub metering 3 is almost equal to that for univariate global active power as shown in Figures 85 – 89 and 90 – 94 respectively. This is because sub metering 3 and global active

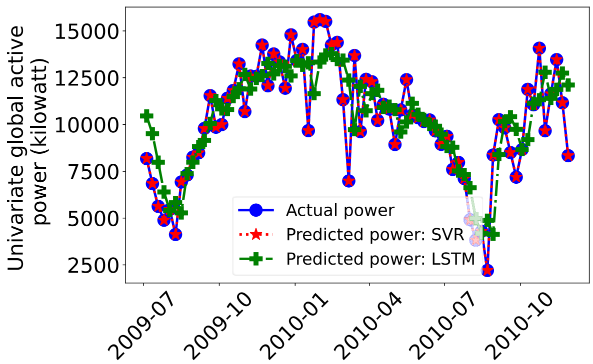

FIGURE 92. Univariate global active power consumption forecast using SVR and LSTM for *$IHEPC_{week}$* dataset.

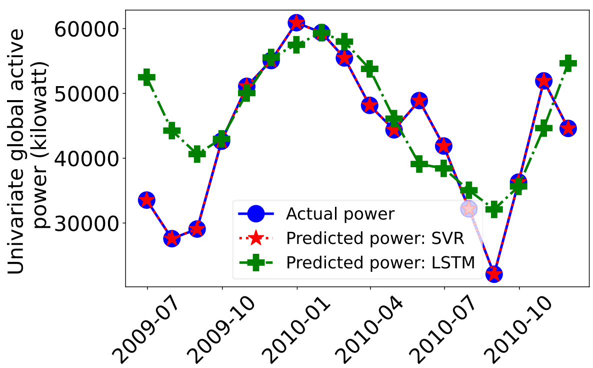

FIGURE 93. Univariate global active power consumption forecast using SVR and LSTM for *$IHEPC_{month}$* dataset.

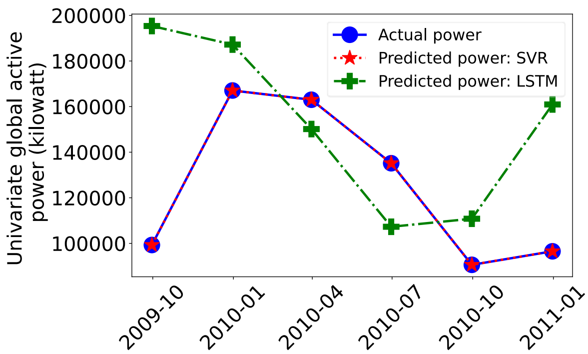

FIGURE 94. Univariate global active power consumption forecast using SVR and LSTM for *$IHEPC_{quarter}$* dataset.

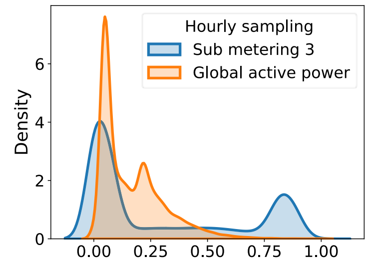

FIGURE 95. Distribution of sub metering 3 and global active power for *$IHEPC_{hour}$* dataset.

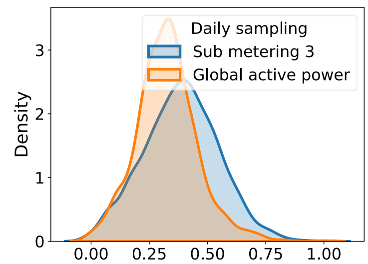

FIGURE 96. Distribution of sub metering 3 and global active power for *$IHEPC_{day}$* dataset.

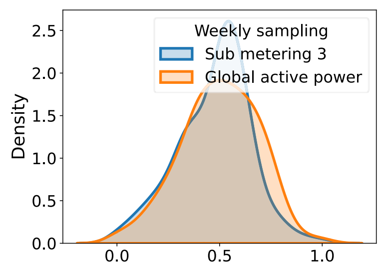

FIGURE 97. Distribution of sub metering 3 and global active power for *$IHEPC_{week}$* dataset.

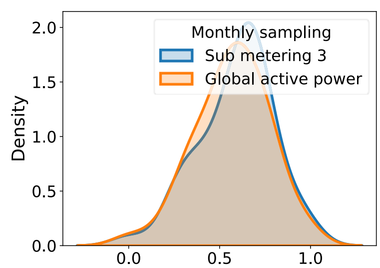

FIGURE 98. Distribution of sub metering 3 and global active power for *$IHEPC_{month}$* dataset.

power are highly correlated as depicted in Figure 3. This is also confirmed by the similar spatial distribution of sub metering 3 data and global active power data for different forecast levels as shown in Figures 95 –99.

Figures 100 – 104 present the multivariate global active power prediction using SVR and LSTM for $IHEPC_{hour}$, $IHEPC_{day}$, $IHEPC_{week}$, $IHEPC_{month}$, and $IHEPC_{quarter}$ validation datasets respectively. As depicted in the figures, SVR outperforms LSTM as it can capture the non-linear data trend for global active power (Figure 11). Comparing the performances of SVR and LSTM for univariate global active power prediction (Figures 90 –94) and multivariate global active power prediction (Figures 100 –104), both the algorithms provide more accurate predictions for multivariate time series. This is because multivariate prediction considers the relationship between all dataset features (i.e., global reactive power, voltage, global intensity, sub metering 1, sub

TABLE 9. Comparison between SVR and LSTM forecast models for the IHECP dataset.

	Sampling									
	Hourly		Daily		Weekly		Monthly		Quarterly	
	Root Mean Square Error (RMSE)									
	SVR	LSTM	SVR	LSTM	SVR	LSTM	SVR	LSTM	SVR	LSTM
Sub metering 1	0.10	0.10	0.10	0.10	0.10	176.76	1442.85	4646.16	18848.15	50089.44
Sub metering 2	46.61	0.10	0.10	0.10	0.10	187.41	1723.28	4601.39	14402.07	38589.48
Sub metering 3	0.10	199.86	0.10	0.10	0.10	335.44	3219.66	16168.95	65185.71	305959.60
Univariate global active power	23.20	0.10	0.10	0.10	0.10	33.61	384.84	1831.04	8148.07	50218.00
Multivariate global active power	0.10	51.36	0.10	0.10	0.10	16.26	288.57	1534.45	8120.64	27980.98
	Symmetric Mean Absolute Percentage Error (sMAPE)									
Sub metering 1	1.73	0.53	0.05	0.00	0.00	1.89	0.90	0.44	0.35	1.73
Sub metering 2	0.70	0.03	0.00	0.00	0.00	1.14	0.81	0.34	0.25	0.70
Sub metering 3	0.09	0.03	0.00	0.00	0.00	0.86	0.28	0.20	0.20	0.09
Univariate global active power	0.18	0.03	0.00	0.00	0.00	0.46	0.22	0.15	0.15	0.18
Multivariate global active power	0.05	0.04	0.00	0.00	0.00	0.28	0.18	0.14	0.14	0.05
	Jensen-Shannon divergence									
Sub metering 1	0.00	0.39	0.00	0.35	0.00	0.17	0.00	0.14	0.00	0.11
Sub metering 2	0.08	0.33	0.00	0.32	0.00	0.15	0.00	0.08	0.00	0.08
Sub metering 3	0.00	0.27	0.00	0.12	0.00	0.09	0.00	0.08	0.00	0.13
Univariate global active power	0.08	0.12	0.00	0.10	0.00	0.07	0.00	0.06	0.00	0.11
Multivariate global active power	0.00	0.05	0.02	0.07	0.00	0.06	0.00	0.07	0.00	0.07

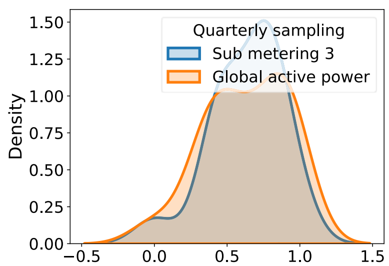

FIGURE 99. Distribution of sub metering 3 and global active power for *$IHEPC_{quarter}$* dataset.

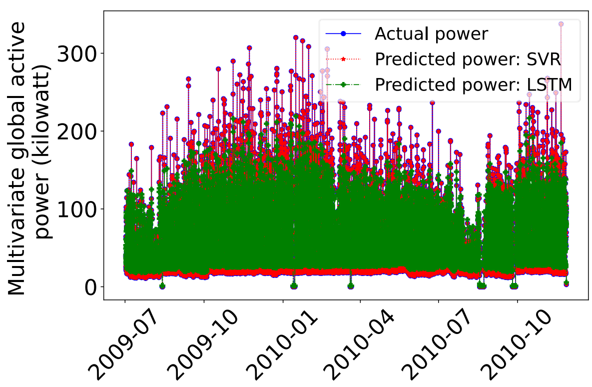

FIGURE 100. Multivariate global active power consumption forecast using SVR and LSTM for *$IHEPC_{hour}$* dataset.

metering 2, and sub metering 3) and global active power, leading to more accurate predictions.

Figures 105 – 108 present the appliances' energy consumption prediction using SVR and LSTM for AEP, AEP_{hour}, AEP_{day}, and AEP_{week} validation datasets

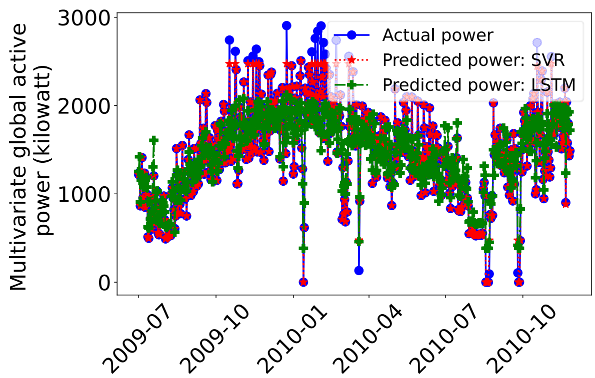

FIGURE 101. Multivariate global active power consumption forecast using SVR and LSTM for *$IHEPC_{day}$* dataset.

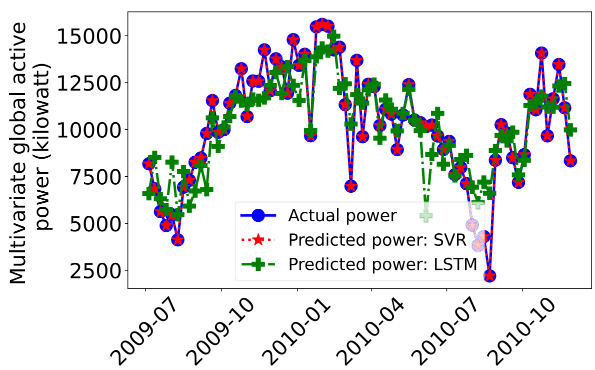

FIGURE 102. Multivariate global active power consumption forecast using SVR and LSTM for *$IHEPC_{week}$* dataset.

respectively. As revealed, overall SVR outperforms LSTM by accurately predicting the energy consumption. This is

TABLE 10. Comparison between SVR and LSTM forecast models for the AEP dataset.

	Sampling							
	Minutely		Hourly		Daily		Weekly	
	SVR	LSTM	SVR	LSTM	SVR	LSTM	SVR	LSTM
Root Mean Square Error (RMSE)	0.09	32.27	0.10	246.62	0.10	4809.80	0.10	12439.23
Symmetric Mean Absolute Percentage Error (sMAPE)	0.00	0.23	0.00	0.28	0.00	0.27	0.00	0.10
Jensen-Shannon divergence	0.00	0.08	0.00	0.12	0.00	0.10	0.00	0.03

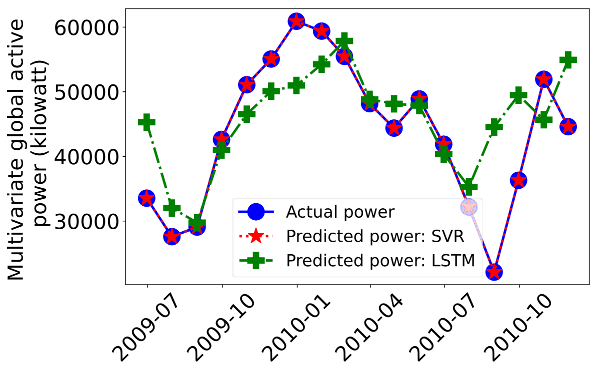

FIGURE 103. Multivariate global active power consumption forecast using SVR and LSTM for $IHEPC_{month}$ dataset.

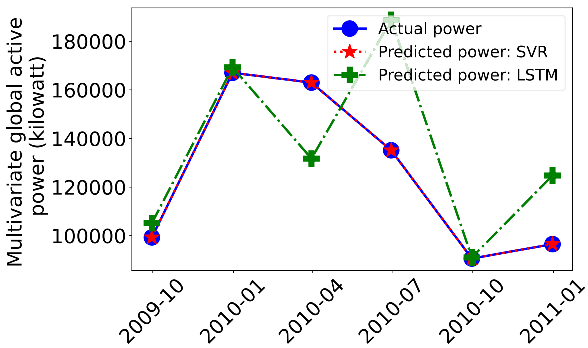

FIGURE 104. Multivariate global active power consumption forecast using SVR and LSTM for $IHEPC_{quarter}$ dataset.

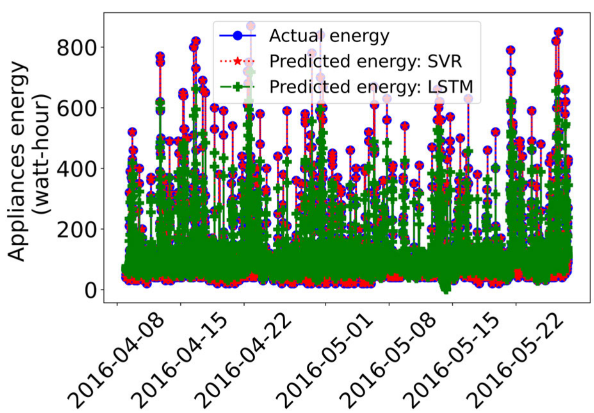

FIGURE 105. Multivariate appliances' energy consumption forecast using SVR and LSTM for *AEP* dataset.

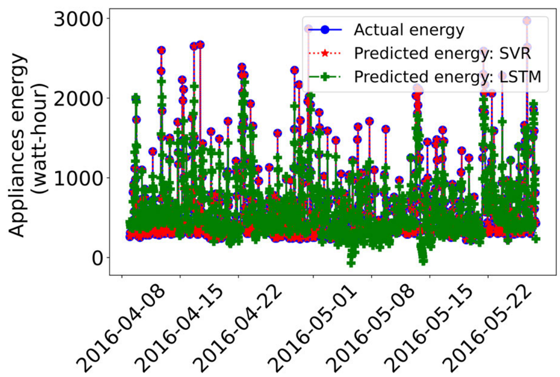

FIGURE 106. Multivariate appliances' energy consumption forecast using SVR and LSTM for AEP_{hour} dataset.

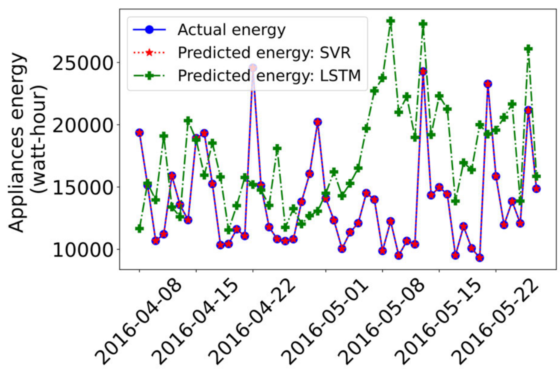

FIGURE 107. Multivariate appliances' energy consumption forecast using SVR and LSTM for AEP_{day} dataset.

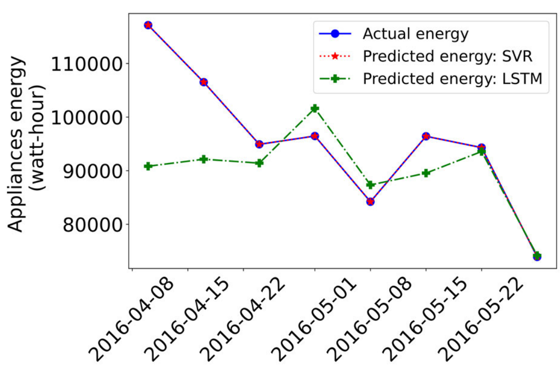

FIGURE 108. Multivariate appliances' energy consumption forecast using SVR and LSTM for AEP_{week} dataset.

because SVR when used with RBF kernel captures the non-linear relationship between appliances' energy and other variables, such as temperatures, humidities, pressure, wind-speed, and visibility. Tables 9 and 10 present the RMSE, sMAPE, and Jensen-Shannon divergence values for the IHEPC and AEP datasets respectively, obtained by SVR and LSTM for different forecast levels. They show that the SVR models outperform the LSTM models for each forecast level.

VI. CONCLUSION AND FUTURE WORK

Residential energy consumption is increasing at an alarming rate due to several factors, such as a growing population, remote work from home post-pandemic, extreme climatic changes, and economic development. High consumption produces a huge amount of carbon dioxide and other GHGs, causing global warming. Consequently, it becomes important to consume energy efficiently. High energy efficiency and reduced GHG emissions can be achieved by accurately forecasting household load and accordingly planning energy generation and usage. Electricity load forecasting at different levels, i.e., hourly, daily, weekly, monthly, and quarterly, can aid energy companies in effectively and efficiently reducing blackouts and planning production, tests, maintenance schedules, investments, constructions, and environmental policies. On the other hand, unit-wise and global load prediction can give household owners insight into the most energy-consuming units/appliances that should be considered for energy savings. Reducing residential energy consumption can help both companies and household owners earn carbon credits, an emerging initiative launched by the Kyoto Protocol, an international agreement linked to the United Nations Framework Convention on Climate Change to reduce emissions of GHGs.

In this paper, we explained the correlation of different energy-consuming units (i.e., kitchen area, laundry area, water heater, and air conditioning) with global household power consumption. Furthermore, we provided insights into the most energy-consuming units based on temporal distribution. Later, we implemented the most used SVR and LSTM models for forecasting electricity load at different levels using univariate and multivariate time series data. We evaluate the prediction performance of sub metering and global household load for different forecasting levels. The dataset used in this study is the largest publicly available household energy consumption dataset. Our experimental result reveals that SVR outperforms LSTM for all forecast levels for both univariate and multivariate data.

For future research work, a larger spectrum of forecast models will be evaluated. In addition, more datasets will be considered that involve appliance-level load forecasting. This will allow household owners to manage energy consumption more efficiently. Furthermore, lightweight forecast models will be implemented in mobile phones for remote energy management in sustainable smart homes.

REFERENCES

[1] International Energy Agency. *Worldwide Total Electricity Consumption*. Accessed: Jul. 24, 2023. [Online]. Available: https://www.iea.org/reports/electricity-information-overview/electricity-consumption

[2] Eurostat. *Energy Consumption in European Union (EU) Households*. Accessed: Jul. 24, 2023. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_consumption_in_households

[3] *Electricity Use of Residential Buildings in Australia*. Accessed: Jul. 24, 2023. [Online]. Available: https://www.energy.gov.au/government-priorities/buildings/residential-buildings

[4] P. Jadwiszczak, J. Jurasz, B. Kaźmierczak, E. Niemierka, and W. Zheng, "Factors shaping A/W heat pumps CO_2 emissions—Evidence from Poland," *Energies*, vol. 14, no. 6, p. 1576, Mar. 2021, doi: 10.3390/en14061576.

[5] M. Alkhraijah, M. Alowaifeer, M. Alsaleh, A. Alfaris, and D. K. Molzahn, "The effects of social distancing on electricity demand considering temperature dependency," *Energies*, vol. 14, no. 2, p. 473, Jan. 2021, doi: 10.3390/en14020473.

[6] P. A. Schirmer, C. Geiger, and I. Mporas, "Residential energy consumption prediction using inter-household energy data and socioeconomic information," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Jan. 2021, pp. 1595–1599, doi: 10.23919/Eusipco47968.2020.9287395.

[7] R. Sendra-Arranz and A. Gutiérrez, "A long short-term memory artificial neural network to predict daily HVAC consumption in buildings," *Energy Buildings*, vol. 216, Jun. 2020, Art. no. 109952, doi: 10.1016/j.enbuild.2020.109952.

[8] M. Aydinalp, V. Ismet Ugursal, and A. S. Fung, "Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks," *Appl. Energy*, vol. 71, no. 2, pp. 87–110, Feb. 2002, doi: 10.1016/s0306-2619(01)00049-6.

[9] C. Bohringer, "The Kyoto protocol: A review and perspectives," *Oxford Rev. Econ. Policy*, vol. 19, no. 3, pp. 451–466, Sep. 2003, doi: 10.1093/oxrep/19.3.451.

[10] S. Demiralay, H. G. Gencer, and S. Bayraci, "Carbon credit futures as an emerging asset: Hedging, diversification and downside risks," *Energy Econ.*, vol. 113, Sep. 2022, Art. no. 106196, doi: 10.1016/j.eneco.2022.106196.

[11] J. Yu and M. L. Mallory, "Exchange rate effect on carbon credit price via energy markets," *J. Int. Money Finance*, vol. 47, pp. 145–161, Oct. 2014, doi: 10.1016/j.jimonfin.2014.04.010.

[12] J. Henzel, Ł. Wróbel, M. Fice, and M. Sikora, "Energy consumption forecasting for the digital-twin model of the building," *Energies*, vol. 15, no. 12, p. 4318, Jun. 2022, doi: 10.3390/en15124318.

[13] J.-S. Chou and N.-T. Ngo, "Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns," *Appl. Energy*, vol. 177, pp. 751–770, Sep. 2016, doi: 10.1016/j.apenergy.2016.05.074.

[14] N. Sakib, E. Hossain, and S. I. Ahamed, "A qualitative study on the united states Internet of Energy: A step towards computational sustainability," *IEEE Access*, vol. 8, pp. 69003–69037, 2020, doi: 10.1109/ACCESS.2020.2986317.

[15] S. Araya, N. Rakesh, and M. Kaur, "Smart home load analysis and LSTM-based short-term load forecasting," in *Proc. Int. Conf. Innov. Inf. Commun. Technol. (IICT)*, Delhi, India, 2020, pp. 123–131, doi: 10.1007/978-3-030-66218-9_14.

[16] M. Alamaniotis, A. Ikonomopoulos, and L. H. Tsoukalas, "Evolutionary multiobjective optimization of kernel-based very-short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1477–1484, Aug. 2012, doi: 10.1109/TPWRS.2012.2184308.

[17] E. Mele, "A review of machine learning algorithms used for load forecasting at microgrid level," in *Proc. Int. Sci. Conf.*, 2019, pp. 452–458, doi: 10.15308/sinteza-2019-452-458.

[18] D. R. Brillinger, *Time Series: Data Analysis and Theory*. Philadelphia, PA, USA: SIAM, 2001, doi: 10.1137/1.9780898719246.

[19] L. Ismail, H. Materwala, and A. Hennebelle, "Forecasting COVID-19 infections in Gulf cooperation council (GCC) countries using machine learning," in *Proc. 13th Int. Conf. Comput. Modeling Simulation*, Jun. 2021, pp. 231–236, doi: 10.1145/3474963.3475844.

[20] R. F. Berriel, A. T. Lopes, A. Rodrigues, F. M. Varejão, and T. Oliveira-Santos, "Monthly energy consumption forecast: A deep learning approach," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 4283–4290, doi: 10.1109/IJCNN.2017.7966398.

[21] L. M. Candanedo, V. Feldheim, and D. Deramaix, "Data driven prediction models of energy use of appliances in a low-energy house," *Energy Buildings*, vol. 140, pp. 81–97, Apr. 2017, doi: 10.1016/j.enbuild.2017.01.083.

[22] Z. Zheng, H. Chen, and X. Luo, "A Kalman filter-based bottom-up approach for household short-term load forecast," *Appl. Energy*, vol. 250, pp. 882–894, Sep. 2019, doi: 10.1016/j.apenergy.2019.05.102.

[23] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, ''Deep learning for estimating building energy consumption,'' *Sustain. Energy, Grids Netw.*, vol. 6, pp. 91–99, Jun. 2016, doi: 10.1016/j.segan.2016.02.005.
[24] N. Jin, F. Yang, Y. Mo, Y. Zeng, X. Zhou, K. Yan, and X. Ma, ''Highly accurate energy consumption forecasting model based on parallel LSTM neural networks,'' *Adv. Eng. Informat.*, vol. 51, Jan. 2022, Art. no. 101442, doi: 10.1016/j.aei.2021.101442.
[25] T. Le, M. T. Vo, T. Kieu, E. Hwang, S. Rho, and S. W. Baik, ''Multiple electric energy consumption forecasting using a cluster-based strategy for transfer learning in smart building,'' *Sensors*, vol. 20, no. 9, p. 2668, May 2020, doi: 10.3390/s20092668.
[26] M. A. Alotaibi, ''Machine learning approach for short-term load forecasting using deep neural network,'' *Energies*, vol. 15, no. 17, p. 6261, Aug. 2022, doi: 10.3390/en15176261.
[27] R. Gonzalez, S. Ahmed, and M. Alamaniotis, ''Implementing very-short-term forecasting of residential load demand using a deep neural network architecture,'' *Energies*, vol. 16, no. 9, p. 3636, Apr. 2023, doi: 10.3390/en16093636.
[28] Y. Yu, X. Si, C. Hu, and J. Zhang, ''A review of recurrent neural networks: LSTM cells and network architectures,'' *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019, doi: 10.1162/neco_a_01199.
[29] C. Cortes and V. Vapnik, ''Support-vector networks,'' *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.
[30] G. Hebrail and A. Berard. *Individual Household Electric Power Consumption Dataset*. Accessed: Jan. 11, 2023. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption
[31] *Smart Meters in London*. Accessed: Jan. 11, 2023. [Online]. Available: https://www.kaggle.com/datasets/jeanmidev/smart-meters-in-london
[32] W. M. Healy. *Net Zero Energy Residential Test Facility Instrumented Data; Year 2*. Accessed: Jan. 11, 2023. [Online]. Available: https://data.nist.gov/od/id/3C53B142D0C3268EE0531A570681EA991497
[33] J. Kelly and W. Knottenbelt, ''The U.K.-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five U.K. homes,'' *Sci. Data*, vol. 2, no. 1, Mar. 2015, Art. no. 150007, doi: 10.1038/sdata.2015.7.
[34] *MagicBox*. Accessed: Jan. 11, 2023. [Online]. Available: http://www.magicbox.etsit.upm.es/
[35] *Survey of Household Energy Use (SEHU)*. Accessed: Jan. 11, 2023. [Online]. Available: https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4403
[36] M. Liu, L. Chen, X. Du, L. Jin, and M. Shang, ''Activated gradients for deep neural networks,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 4, pp. 2156–2168, Apr. 2023, doi: 10.1109/TNNLS.2021.3106044.
[37] J. Benesty, J. Chen, Y. Huang, and I. Cohen, ''Pearson correlation coefficient,'' in *Noise Reduction in Speech Processing*. Springer, 2009, pp. 1–4, doi: 10.1007/978-3-642-00296-0_5.
[38] A. Fernández-Montes, D. Fernández-Cerero, F. Escalera-González, A. Jakóbik, B. Bermejo, and C. Juiz, ''SimilarityTS: Toolkit for the evaluation of similarity for multivariate time series,'' *SoftwareX*, vol. 24, Dec. 2023, Art. no. 101527, doi: 10.1016/j.softx.2023.101527.

LEILA ISMAIL (Member, IEEE) received the joint Ph.D. degree (Hons.) from the Grenoble Institute of Technology and French Institute for Research in Computer Science and Technology (INRIA), France. She is currently the Founding Director of the Intelligent Distributed Computing and Systems (INDUCE) Laboratory, an Associate Professor with the Department of Computer Science and Software Engineering, United Arab Emirates University, and a Visiting Associate Professor with the Cloud Computing and Distributed Systems (CLOUDS) Laboratory, School of Computing and Information Systems, Faculty of Engineering and Information Technology, The University of Melbourne, Australia. She has vast industrial and academic experience with the Sun Microsystems Research and Development Center, France, on the design and implementation of realtime and highly available distributed systems and participated in the deposit of a U.S. patent. She was a Lecturer with the Grenoble Institute of Technology and an Adjunct Professor with the Digital Ecosystems and Business Intelligence Institute, Curtin University, Australia. She is active in international collaborations and publishes her research results in prestigious journals and international conferences. Her research interests include developing efficient and sustainable smart city digital ecosystems for energy savings, smart healthcare, and smart transportation systems, enabled by the IoT, artificial intelligence, machine learning, edge and cloud computing, and blockchain. She is a recipient of several awards and appreciation certificates. She has been invited as a Keynote Speaker to several conferences, including the Women in Data Science International Conference (WiDS 2021), organized by Stanford University, USA. She served as an Associate Editor for the *International Journal of Parallel, Emergent and Distributed Systems* for several years, served as the Chair, the Co-Chair, and the Track Chair for many IEEE and ACM international conferences, including being the General Co-Chair for IEEE DEST 2009, the General Chair, the Technical Program Chair, and the Organizing Committee Chair for the 11th International Conference on Innovations in Information Technology 2015 (IIT'15) for which she got the support of the IEEE Computer Society (HQs, USA) Technical Sponsorship. She is an Editor of the *Information Innovation Technology in Smart Cities* (Nature Springer, 2018). Her research interests include artificial intelligence, machine learning, deep learning, energy savings, sustainable smarty city, the Internet of Things, edge and cloud computing, and blockchain. She is top 2% world's cited scientist in AI, networking and communication, and information and communication technologies by Elsevier and Stanford University, USA.

HUNED MATERWALA received the Bachelor of Technology degree in instrumentation and control engineering from Gujarat Technological University, India, in 2013, and the Master of Technology degree in control systems from the Department of Electrical and Electronics Engineering, Amity University, Uttar Pradesh, India, in 2016. He is currently pursuing the Ph.D. degree with the Intelligent Distributed Computing and Systems (INDUCE) Laboratory, College of Information Technology, United Arab Emirates University, United Arab Emirates. He is also a Research Assistant with the INDUCE Laboratory, College of Information Technology, United Arab Emirates University. His research interests include edge and cloud computing, energy efficiency, distributed systems, smart transportation, smart healthcare, and sustainable smart city.

FIDA K. DANKAR received the Ph.D. degree in computer science from the University of Ottawa, Canada. She is currently a Visiting Associate Professor with New York University Abu Dhabi (NYUAD). Prior to joining NYUAD, she was an Associate Professor with United Arab Emirates (UAE) University. Prior to joining UAE University, she was a Research Scientist with the IBM Canada Research and Development Centre, where she worked on the secure and private mining of distributed health data. Prior to joining IBM, she worked for eight years with the Children's Hospital of Eastern Ontario Research Institute, her work centered around facilitating the sharing of health information for secondary purposes while protecting the privacy of patients and the identity of providers. Her research interests include the development of multidisciplinary approaches for the private and secure mining/sharing of health data that uses methods from cryptography, biomedical knowledge modeling, and policy analysis.